

Who wants the US space station?

The US plan to put a space station into an orbit about the Earth has little to commend it except that it will show that the United States is still in space. But there are better and safer ways of leading.

THE motto that it never rains without pouring must seem apt to the embattled administration of the United States. Very little seems to be going right. That is why, in a week in which the reputation of the presidency is likely to be further damaged by Senator John Tower's report on the sale of munitions to Iran, it may seem unkind to be rubbing salt in another of the administration's self-inflicted wounds. Yet, on the same calculation, this is when the astringent is most likely to be effective. And that is why the United States should now bite on another bullet and abandon its plan to build a space station in an orbit about the Earth.

When first put forward, three years ago, the plan seemed to some a natural way of building on the promised success of the space shuttle, a remarkable technical achievement in spite of last year's tragic accident. Given a means of carrying people and equipment into orbit for a week or so, why not use the same device to establish something more permanent? The outcome was an enterprise, jointly with Canada, the European Space Agency and Japan, to install a space station in 1994. Now it is inevitable that completion must be delayed, not so much because of last year's accident but because of technical difficulties that have increased the estimated cost of the original design from \$8,000 million to more than \$14,000 million. During the next six months, the US National Aeronautics and Space Administration (NASA) will be seeking some kind of compromise with a budget-conscious Congress, leaving its partners in limbo while the while. Both parties to that dialogue should appreciate that it would be better to cancel the project now, before it becomes an albatross around NASA's neck.

The least important facet of the argument that lies ahead is that which has received the most attention in the past few weeks, the possibility that the space station might provide house-room for experiments commissioned by the US military. NASA rightly says that its partners have known from the outset that the US Department of Defense might wish to share the space station with civilian users (some of them commercial and thus as obsessed with secrecy as the military). Present anxiety on this question, most acute in Europe, is largely the result of poor timing. The Pentagon's reasonable request, towards the end of last year, for a say on the agreements being negotiated with potential partners coincided with its push for a decision to deploy some elements of the Strategic Defense Initiative (SDI) long before 1993, the date foreseen when SDI was launched on an astonished world. Logically, there is no connection between these developments, but NASA and the Pentagon have been slow to see the potential connection, the fear of many in Europe that an accommodation on military uses for the space station may seem to imply approval for the early deployment of SDI. The difficulty would not have arisen if the diplomats had done their paperwork more quickly. The next few weeks should tell whether it can be chased away. Failure would weigh heavily with Congress during the budget process.

Other reasons for cancelling the space station bear more directly on the plan itself. One difficulty is that the capacity of the shuttle to carry hardware into orbit will be much less than foreseen three years ago; even when the lost craft is replaced, the frequency of flights will be more like one a month than one a

fortnight. What will happen to the payloads crowded out by the components of the space station? Another problem is the prospect that this project, like the shuttle programme after the first flush of enthusiasm, will have to be run on a shoestring. Although it is too soon to know whether the new US Congress will face up to its fiscal responsibilities to bring the budget deficit within limits, it is more than likely that a necessarily uncertain project such as the space station costing, even now nearly as much as the Apollo programme of the 1960s will, at some point in the next seven years, be forced to rub along with funds insufficient to guarantee success. And what would happen if the linked evils of the US budget and trade deficits should bring about the sharp recession about which economists speculate more openly every day?

Strategy

Prudence, and NASA's painful experience of being so committed to the shuttle at a time of budgetary restraint that it had to leave aside many smaller useful projects, argue for a less risky strategy. Neither the administration nor the US Congress seems to realize that the most curious feature of their strategy in space has been its discontinuity. Having reached the Moon, NASA built the shuttle. Having, with the Pioneer and Voyager space craft, shown how to explore the distant reaches of the Solar System, NASA has no plans to follow through. Having put two superb X-ray telescopes in orbit, NASA seems willing to leave to others the design of their successors. Meanwhile, the agency can only receive in embarrassed silence the repeated reasoned arguments from official and unofficial advisory groups that urge this or that good project in the exploration of space. (Another list of unrequitable wishes from NASA's own science advisory committee is due out this week.) The result has been the alienation of a large part of the scientific community.

Nor can NASA claim that it has no more money for basic science because it is pursuing the applications of satellite technology instead. Here, too, NASA is a feckless innovator. Having pioneered Earth observation satellites, it must contemplate the superior performance of the French instrument called SPOT. The hope that the US Navy would assume (from the National Oceanographic and Atmospheric Agency) responsibility for the sea-surface observation satellites seems to have been thwarted by the Navy's cash problems, but is it really the US government's intention that such projects should be the vicarious casualties of random budget shortages? And while NASA may be right to have decided (in the budget for the next fiscal year) that the project to build an advanced telecommunications satellite can be safely left to private industry, there is now no knowing when that will be done. In applications as in the basic science programme, NASA cuts the figure of a fickle will-o'-the-wisp, trying one thing and then another. The agency's conventional wisdom is that it has no choice, for Congress would 'never' agree to spend much more than \$2,000 million on these unspectacular tasks. The statement would be more telling if NASA had ever sought to persuade people otherwise.

It is thus not surprising that US taxpayers and even NASA officials are beginning to ask how "leadership" in space can be



assured, even recovered. As usually asked, the question is vacuous. Leadership is taken as a synonym for 'in the lead' or 'first'. On this philosophically bankrupt basis, last year's shuttle accident is bound to seem not merely a tragedy but a threat to leadership. The truth is that a better synonym for leadership in technological connections is expertness, the capacity to carry out difficult tasks with ease and confidence. Now that being ubiquitously first (on the Moon, in a space station or even on Mars) is financially and technically impossible, not to say philosophically pointless, should not NASA aim at the more durable leadership that could come from doing excellently what it does well already? It would be a less spectacular programme, but the United States might welcome a respite from sensation. □

Ivan Boesky's friends

The insider-trading scandal seems unstoppable, but there are some things that could be done.

THE great worldwide financial scandal goes from bad to worse, but governments and financial communities appear mesmerized by what has overtaken them, and to be incapable of response. Last week, Mr Dennis Levine, the insider-trader who first blew the whistle on Mr Ivan Boesky, was sentenced by a New York court to four parallel terms of two years' imprisonment. A week earlier, an even bigger fish, Mr Martin A. Siegal, had pleaded guilty to two felony charges and had declared that he is cooperating with the authorities (which no doubt signifies that there are more surprises to come). By all accounts, the indictment of Siegal explains that a courier working for Boesky had at one stage handed him by prearrangement in a hotel lobby a suitcase containing \$150,000 in notes; the money was used in succeeding years to pay minor household bills. A few days earlier, a senior member of the stockbroking firm of Goldman, Sachs and Co. was arrested together with two others from Siegal's old firm, Kidder Peabody (but Goldman, Sachs says it will pay the cost of defending its Mr Robert M. Freeman).

These events would be fascinating if they were not also tragic and frightening. Commenting on the Siegal case last week, the *Wall Street Journal* says that it shows that insider-trading has become "systemic" in New York. That is frightening because it gives the lie to the belief, much cultivated by the self-interested financial communities, that the markets are efficient and are thus reliable ways of deciding the fate of enterprises of all kinds, from established industries to struggling new ventures. To the extent that the galloping scandal (and its Guinness offshoot in London) shows that it is possible to manipulate the price of publicly traded stock, even at some cost, it is no longer possible to believe that what happens on the markets equates with common sense. Worse still, the manipulators have been driven by the attraction of personal gain, in Boesky's case on a scale that enabled him to pay a fine of \$100 million and still have some left over; in doing so, they have been stealing from other people, often those ordinary mortals who put their savings in pension funds and other financial securities.

What can be done? It remains to be seen what are the effects of the gaol sentences now being handed out, but of necessity they will not be permanent. In the US Congress, Senator William Proxmire has followed the British Labour Party in suggesting tighter regulation of the takeover process (which attends to the substrate of corruption, not its cause). In both places (and elsewhere) it would make better sense if it were a legal requirement that the true owners of all traded stocks should be instantaneously and publicly known. Then at least there would be a chance that the professionals would catch each other out. And why should not the people who work in the financial markets be forbidden from owning stock on their own account? If it should be said that people would not do this job if deprived of the opportunity for making a little extra on the side, that would be a proof that insider-trading is indeed systemic. □

US university risks

British universities have lost their political friends. Will US universities follow suit?

COULD the university system in the United States be running headlong into the problems that have overtaken British universities in the past decade? On the face of things, that is impossible. Unlike British universities, whose costs are mostly met directly from the public purse, universities in the United States are either private institutions whose autonomy is beyond dispute or are supported by state governments jealous of their institutions' achievements and conscious of their economic importance. That is the theory. But US universities, and especially the private institutions, may now be courting just the danger that, ten years ago, marked the onset of the British government's beastliness to its universities. The US universities, or some of them, run the risk of losing their political support, ironically even when the economic value of the skills that universities can engender is more keenly appreciated than ever.

The underlying issue, first in Britain and now in the United States, is public accountability. In Britain in 1967, Mrs Shirley Williams, then Secretary of State for Education and Science in a Labour government, asked the universities two pointed questions: how would they respond to the demographic prospect of falling enrolments, and what suggestions did they have for helping a financially beleaguered government contain the cost of student maintenance? The replies — that shrinking age-groups do not imply shrinking universities, and that there is a case (in British circumstances) for paying maintenance costs to all, are reasonable points of view — but took no account of the administrative reasons why even a sympathetic minister should have been compelled to raise these awkward questions.

Nobody would suggest that present circumstances in the United States are the same. Mr William J. Bennett, the Secretary of Education, makes very few noises of the kind that academics like to hear (see p.751). Moreover, federal aid to higher education is a small percentage of the total cost of higher education, and shrinking. Yet Bennett, like Williams, is asking questions that universities would be well advised to answer patiently. Is it, for example, true that tuition fees have increased on the average by 58 per cent in the past five years, twice as quickly as inflation, and if so why? If, as some institutions will be tempted to say, the explanation is that higher tuition fees are needed to meet the disproportionately increased need for student aid, there will be many who ask what business it is of universities to operate a redistributive taxation system. And is it true that average university salaries have kept pace with inflation in the past decade, when one of the most striking social changes in the United States has been a relative decline of individual, as distinct from family, incomes (as women go out to work)? Universities will say that to pursue excellence they must compete for what talent there is, but there are many who will protest that the bidding process differs from that among baseball teams in that the costs fall on students who have no choice but to pay up.

What US universities may not have understood clearly is that Mr Bennett's questions, even if misconceived, may anger those who hear them answered superficially, disdainfully or casuistically. What the British experience has shown is that it is easier to provoke a hostile animus than to mollify it, especially among politicians. The moral is that Bennett should be given a reasoned and comprehensive answer to the questions he keeps flinging at the system, but also a more constructive account of what needs to be done to keep the system healthy. The weakest point in his argument is the assumption that even students from the poorest families will become so aware of the benefits of higher education that they will shoulder the burdens of commercially priced loans to win the rewards. It is unlikely that the United States is any better able than Britain to risk the loss of talent provoked in that way. □

World trade war looms over microchip accord

- USA demands that Japan fulfil trade pact
- EEC calls for ruling on 'illegal' trade plan

London and Washington

A MAJOR trade row is about to erupt, embroiling the governments of the United States, Japan and Europe. At the heart of the conflict is frustration in the United States at unsuccessful attempts both to curb the rising tide of Japanese microchips and to tighten its grip on its own components sold overseas.

The fears of the US semiconductor industry, which has accused the Japanese of 'dumping' microchips below their market value, had been partially allayed by an agreement signed with Japan last summer. But the Americans have not been able to satisfy themselves that the Japanese are adhering to the terms of the agreement — increasing the price of microchips while assisting US companies to penetrate the Japanese market — and at the same time have provoked the wrath of European allies who consider the pact a cartel.

The United States is now forced into trying other methods of protecting its home electronics industries and preventing military programmes from depending on imported products and technology because of a weakened infrastructure.

For its part, the European Economic Community (EEC) is making plans to take the US-Japanese microchip accord to a meeting of the General Agreement of Tariffs and Trade (GATT) next week to seek a ruling on the legality of the pact. It says the agreement is a breach of international trading procedures.

Meanwhile, a Defense Science Board Task Force report looking into dependence of foreign semiconductors for military hardware says that unless something is done quickly, the US semiconductor industry will die. The report urges the Pentagon to pump more than \$1,000 million into a new institute for advanced, high-yield manufacturing techniques that will restore US self-sufficiency.

This concept differs from a similar industry-sponsored programme called Sematech because it focuses on generic process technology rather than large-scale production. The task force also encourages the Defense Department to spend \$50 million per year to establish eight university centres of excellence in semiconductor science and engineering.

The report focuses on dynamic, random access memories (DRAMs). US manufacturers used to produce nearly all of the world's supply, but now they hold less than 5 per cent of the world market. Japan

has taken over the lead in DRAMs, and is moving up in other areas of the industry.

Part of Japan's success comes from aggressive marketing and government support and protection for domestic industries. But the key, according to the task force report, is Japan's superior manufacturing practices. Japanese manufacturers have invested approximately 35 per cent of sales in plants and equipment, compared with 20 per cent for their US counterparts. Japanese firms also outspend US firms in research — 13 per cent compared with 10 per cent of sales. US research funds are spent on developing products for immediate sale; in Japan more is spent on long-term improvements.

To solve the problem, the report recommends that US industry be permitted to form a consortium called the Manufacturing Institute for Semiconductors. The Pentagon would provide a \$250 million capital investment, and \$200 million per year in contract support for five years. The consortium would not only assure US military needs, but would also help restore commercial industry. The report suggests that the institute's efforts be focused on advanced manufacturing techniques for the next generation of DRAMs, specifically the 64-megabit DRAM.

The Semiconductor Industry Association was quick to endorse the task force report. The association will decide on its plans for the future of Sematech at a board meeting in Washington next week. But for the present, the industry is still struggling.

The trade agreement with Japan worked out last summer has not brought hoped-for relief to the US industry. Association officials have accused Japanese manufacturers of violating agreements to stop dumping chips — selling them below their manufacturing costs. The industry has asked the government to impose financial penalties on Japanese manufacturers for these actions, and take steps to open Japanese markets for US chips.

On another front, Britain, like other European partners, has been questioning the right of the United States to demand that no 'American' technology be re-exported from Europe without the permission of the Department of Commerce. The issue came to the fore again last week with Mr Paul Channon, the UK Secretary of State for Trade and Industry, rejecting a US request for audits on British companies to monitor their behaviour.

Bill Johnstone & Joseph Palca

UK space plan for government go-ahead at last

London

THIS week, a cabinet meeting, expected to be chaired by Prime Minister Mrs Margaret Thatcher, will try to arrive at a decision on the future of Britain's space research and investment. A 15-year plan, drawn up by the British National Space Centre (BNSC), which has called for about £200 million a year, is to be discussed with ministers, with particular attention to be paid to the views of the ministries that might suffer if money were diverted to a space programme.

The Treasury is likely to be the most formidable opponent to the plan and is unhappy about any laxity in the Public Sector Borrowing Requirement (PSBR), despite disclosures last week that the government's finances are well in the black.

Thatcher could be asked to be the final arbitrator, as some government ministers are prepared to side with the Treasury if the alternative is reduced budgets. The space research and investment funds at present come largely from the Departments of Trade and Industry and Environment and the Ministry of Defence.

The BNSC plan requires not only the doubling of UK investment in space but also that it should control those diverse funds. That aspect of the plan is expected to be opposed by the respective departments and some of the research councils.

The BNSC had to postpone allocating industrial development contracts last month because of a lack of funds caused by the government's vacillations, and the centre was becoming increasingly embarrassed by its inability to commit Britain to the new projects to be undertaken by the European Space Agency (ESA), of which the United Kingdom is a member. BNSC officials have been meeting their European counterparts weekly since the beginning of the year, preparing for a full ministerial council meeting of ESA in June, but cannot outline Britain's space plans.

The government has been extremely reluctant to invest in research or industry if such funding can be acquired from the private sector. The BNSC, however, says that the government must take a lead in developing technologies and in research that shows no obvious commercial return.

The scientific advisers to the Cabinet Office prefer a more modest approach to Britain's space ambitions. The proposed BNSC budget is a commitment not favoured by the advisers. A plan in which spending would approach an annual target budget over a three-year period has more popular appeal.

Bill Johnstone

US call to end CFC emissions

Washington

EUROPE will this week be under strong pressure from the United States to take steps to control the discharge of chlorofluorocarbons (CFCs) into the atmosphere. The occasion will be a meeting in Vienna, starting 23 February, of the Vienna Convention on the Protection of the Ozone Layer, of which the European Economic Community (EEC) is a member.

The US delegation is calling for a freeze on the emission of CFCs, followed by a long-term planned reduction of emissions by as much as 95 per cent. This represents a considerable strengthening of the US position since the last round of negotiations in Geneva in December, when the participants got no further than deciding that something must be done.

Two bills introduced in the US Senate last week are designed to support the US negotiations. Both threaten to implement trade restrictions against countries that do not virtually cease production of ozone-depleting chemicals — brominated hydrocarbons as well as CFCs. Under the terms of the bills, US production will be decreased over a 6–8 year period to a ceiling of 5 per cent of present production. The chemicals are used in refrigerants, solvents and foam packaging for food; the use of CFCs in aerosol propellants was banned in the United States in 1978 and afterwards in Switzerland, Canada and several Scandinavian countries.

The sponsors of the Senate bills, Senators Max Baucus and John Chafee, cite a prediction of the Environmental Protection Agency that there would be 40 million excess skin cancers, resulting in 800,000 deaths in the United States alone over the next 80 years, if nothing is done to curb CFC emissions.

The situation is complicated by disagreements within the EEC about the seriousness of the problem. According to one Senate staff member, the United Kingdom and France are "not willing to admit there is a problem" of ozone depletion, while others did not attend the last round of negotiations. On the other hand, Denmark (which has no CFC industry to protect), has already banned the use of CFCs in aerosols — and may find other EEC nations taking action against it for restraint of trade.

The official US negotiating position in Vienna is less aggressive than the Senate bills. Ambassador Richard Benedick, leader of the US team, hinted last week that the United States would consider signing agreements with individual countries if no general agreement was reached.

Steven Dickman

Row takes new turn over US plagiarism of Soviet books

London

THE case of Dr Hollis Chen, the Ohio University professor accused of plagiarism (see *Nature* 324, 198; 1986), has taken a new turn. Chen was accused, in the pages of the Moscow weekly *Literaturnaya Gazeta*, of having pirated two books by a member of the Byelorussian Academy of Sciences, incorporating them into his textbook *Theory of Electromagnetic Waves*. The attack was signed by the president of the Byelorussian Academy, Dr Mikolaj Barysevic, and a three-times Stale Prize winner, Dr Barys Sciapanau; Chen's defence was that he had never read the works in question, that he speaks no Russian, and that in any case he was writing a university textbook, not a scholarly monograph. He made no claim to originality, he said, and was not responsible for the dust-jacket blurb which suggested that this was the first time the subject was treated from a "coordinate-free" approach.

The second accusation, published in the *Literaturnaya Gazeta* on 28 January, and signed by two other Byelorussian scientists, raises a new charge. It attacks Chen's claim that he did not know the books, stating that a copy of one of the disputed works, the *Theory of Gyrotropy*, was mailed to him by the author, Dr Fiodar Fiodarau. Better known in the West under the Russian version of his name, Fedorov, Fiodarau maintains that in January 1978, Chen wrote to the *Navuka i Technika* press in Minsk to order a book entitled *The Theory of Optical Activity* of which Fiodarau was a co-author. This book, although advertised, had never been published and was superseded by the *Theory of Gyrotropy*. Accordingly, the publishers forwarded Chen's letter to Fiodarau, who replied by sending him a copy of the latter work, says Fiodarau. Chen admits ordering *The Theory of Optical Activity*, but denies ever receiving any acknowledgement of his letter, let alone a book.

Following the earlier article in *Nature* and a similar article in the *New York Times* by Dr Theodore Shabad of Columbia University, the University of Ohio appointed a special committee to look into the charges. The committee was composed of the head of the department of physics and astronomy, Dr Edward Sanford, and two visiting professors, Dr Clayborne D. Taylor of Mississippi State University, and Dr John M. Alexander of the University of Wales.

Their findings, in effect, deny the charges that Chen had copied four chapters almost completely, and that his diagrams and some 1,000 equations were copied unchanged. They conclude that "the test of plagiarism in the case of

teaching material of the type found in textbooks, therefore, as distinct from the possible plagiarism of some completely innovative theory of technique, must be a subjective one, based on the degree and extent of the similarity involved". Chen's work, they said, uses a more modern mathematical treatment, and contains far more diagrams.

Nevertheless, the committee noted some "significant similarities", although not to the extent alleged in the Soviet charges. "Involved are two sections totalling about seven pages in length which are separated by completely disparate developments of the theories involved", they say, and while clearing Chen of the charges of extensive plagiarism, they allow that "there may possibly exist some grounds to believe that Chen made use of a small amount of introductory material from a secondary source" (which itself may have derived from Fiodarau).

The committee met before the appearance of the second *Literaturnaya Gazeta* article, and at that time, it seems, Chen made no mention of his 1978 letter to the Byelorussian publishers. The committee accordingly met again last Friday to discuss the second batch of charges, and again decided in Chen's favour.

So far, the committee has had no chance to see the *Theory of Gyrotropy*, which is not readily available in the West. One wonders, indeed, whether the cause of academic enquiry might not have been better served had the Byelorussian Academy made a discreet approach to Ohio University, submitting copies of the Fiodarau books and marking the disputed passages, rather than resorting to the Soviet press and then, via the Novosti Agency in London and the *Literaturnaya Gazeta* correspondent in New York, ensuring that western journalists noticed the allegations.

One would have preferred, too, a more academic and less polemical approach from the Soviet side — the second *Literaturnaya Gazeta* referred to this correspondent as "a certain" Vera Rich, cast doubts on her professional competence, hinted that *Nature's* response was too rapid, and increased the number of allegedly pirated chapters from four to five. The University of Ohio, on the other hand, might have been well-advised, instead of an in-house enquiry, to refer the whole matter to some external body, perhaps to one of the many learned societies to which Chen belongs, and from which, the first *Literaturnaya Gazeta* article made clear, the Byelorussian Academy would like to see him expelled.

Vera Rich

International winners for Japan's version of Nobel Prize

Tokyo

WINNERS of the Japan Prize, Japan's answer to the Nobel Prize, were announced last week. This year's prizes go to Dr Theodore H. Maiman, father of the laser, and to two noted agronomists, Drs Henry M. Beachell and Gurdev S. Khush. They will receive their awards on 14 April.

The prize, now in its third year, is worth ¥50 million (US\$330,000) and is awarded in two fields each year by the Science and Technology Foundation of Japan. The foundation was set up in 1983 with a donation of ¥3,000 million from Matsushita Electric Industrial Co. But plummeting interest rates in Japan have eaten into the foundation's earnings and Konosuke Matsushita, founder of the giant electronics company and president of the foundation, recently supplemented the funds with a personal contribution of ¥2,000 million.

The prize is awarded to "persons recognized as having served the cause of peace and prosperity for mankind through original and outstanding achievements in

science and technology", and is oriented towards the applied sciences.

The two fields selected this year are improvements in biological functions and electro-optics. Maiman, a US citizen, won the prize in the latter category for his development of the ruby laser (*Nature* 187, 493-494; 1960), and Beachell and Khush share a prize for their development of superior rice strains, which brought about a 'green revolution' in Asia.

Beachell, a native of Nebraska in the United States, developed the IR8 variety of rice in 1966 at the International Rice Research Institute in the Philippines. The IR8 strain has high photosynthetic efficiency and brought a rapid increase in rice production in tropical and sub-tropical environments. Khush, from India, succeeded Beachell in 1972 as head of the institute's plant breeding department and through crossbreeding of 13 parent plants from six countries developed the IR36 strain, which, as well as having high yield, excels in resistance to disease, insects and adverse soil conditions. **David Swinbanks**



Gurdev S. Khush, Henry M. Beachell and Theodore H. Maiman

Indian technology missions told to hurry up

Bangalore

MR Rajiv Gandhi, the Indian Prime Minister, is dissatisfied at the tardy progress of the technology missions meant to provide safe drinking water, better communications, universal immunization and vaccination by 1990 and increasing oil seed production. He has therefore called for a more urgent approach from his ministers and he has directed the ministries and departments involved in programme implementation to keep him regularly formed of progressed. At a review meeting of the missions held early this year, Gandhi made it plain that "ministers and bureaucrats are equally responsible for the successful implementation of these missions and I will not spare anybody".

He went on to say that the objectives of each mission must be quickly achieved. "The objectives have to be specific, tangible and measurable. It is imperative that specific time-bound tasks are spelled out

clearly and allotted to specific institutions for implementation."

Eleven pilot projects are under way in the drinking water mission in 15,000 villages. The communications mission is chiefly aimed at increasing the efficiency of telegram delivery in 500 towns and cities and the target for the telephone network is to bring down the fault rate from 35 per 100 telephones to 10 by 1990.

The technology mission in vaccination and immunization plans to cover 92 million pregnant women and 82 million infants by 1990. The chief diseases to be covered under this programme are polio, measles, tuberculosis, diphtheria, tetanus and typhoid. In oil seeds, the mission is hoping for an output of 24-26 million tonnes by the year 2000, to which end work on development of high yielding varieties of groundnut, rapeseed, mustards, soyabean and sunflower has been intensified.

Radhakrishna Rao

Regulated misconduct

Washington

THE National Science Foundation (NSF) plans to adopt formal rules for the regulation of misconduct by recipients of its grants. The rules, formally approved by a meeting of the National Science Board last week and published in draft form in the *Federal Register* for 10 February, are essentially the same as those in force at the National Institutes of Health. Misconduct is taken to include fabrication, falsification and plagiarism or similar deviations, failure to comply with regulations for the ethical treatment of human subjects and animals, and the breaking of laws applicable to research. According to Erich Bloch, director of NSF, only two cases of misconduct of this kind have come to light among recipients of NSF grants. Under the new procedures, which are open for comment until 13 April, first responsibility for investigating allegations of misconduct will lie with the recipient institution, in the light of which NSF will decide whether to hold an investigation of its own. There is a formal mechanism for appeal. Anonymous informants will be dealt with confidentially. Penalties range from a reprimand to the cancellation of existing grants and the declaration that individuals or even institutions are ineligible for future grants. □

Canadian Advisory Board

Ottawa

CANADA's latest reorganization of research administration was signalled last week when the Prime Minister, Brian Mulroney, held the first meeting of the National Advisory Board on Science and Technology, which includes representatives of the Canadian scientific, business, labour and professional communities. The board, reporting directly to the prime minister, is intended as a "house of sober second thought" providing advice on science and technology policy, especially as these affect the international competitiveness of Canada's high-technology industries. Meanwhile, the draft of a comprehensive national science policy, drawn up after consultation with federal, provincial and territorial ministers, is being revised for publication in the early spring. The intention is that it should be the foundation of a "national policy action plan" to be worked out in the next two years. **C.E.**

New Woods Hole Director

Washington

THE trustees of the Marine Biological Laboratory at Woods Hole have elected Harlyn O. Halvorson of Brandeis University as the laboratory's new director. He will take up the post in August. Halvorson, who has been affiliated with Woods Hole since 1964, replaces J. Richard Whittaker, acting director since the retirement of Paul Weeks in October 1986. **S.D.**

Chiefs dominate indians at an annual science smorgasbord

Chicago

THE gamble by the American Association for the Advancement of Science (AAAS) to hold this year's annual meeting here in mid-winter appears to have succeeded; there has been no snow during the five-day meeting, while the biting wind off Lake Michigan has not been much worse than it can be in summer. The result is that there were roughly 3,400 members (a lot better than last year) at the smorgasbord of hard science made popular (or at least intelligible) and of public policy made technical. There were sessions on everything from particle physics and arms control to the state of Lake Michigan (whose level has risen 1.5 m in five years) and the chance to see night baseball at Wrigley Field, the home of the Chicago Cubs.

The most constant crowd was that at the neuroscience symposium organized by Dr Daniel L. Alkon of the Marine Biological Station at Woods Hole, a kind of current-awareness course equally suitable for outsiders and for those working in the field. And the highlight of that was the astonishing cine-film-footage from Thomas S. Reese (also from Woods Hole) of the movement of mitochondria and packets of neurotransmitter substances along the cytoskeletal microtubules with which the neuronal cytoplasm seems to be packed. It is not so much the measured speed (up to 40 cm a day) as the manner of travel that had Reese's audience spellbound. Mitochondria seem to wriggle their way along the microtubules like loose-fitting railway wagons along a track, sometimes half a dozen or more in line, occasionally changing track.

Other crowds were not so big (AAAS's worry) and seemed sometimes to be outnumbered by the speakers; the effect can be anticlimactic if a panel wishes to pose grand questions best suited to crowded halls, as did the speakers in the five days of talks about the US space programme after Challenger.

The government view is that the manned space station is the "cornerstone" of the next decade's programme for NASA (National Aeronautics and Space Administration), but academic researchers dissent. Ricardo Giacconi, developer of the Einstein X-ray satellite and now director of the NASA-supported Space Telescope Science Institute, flatly told a panel of government officials that there is "very little" support for the space station in the academic community. Giacconi's view is that NASA's commitment to people in space has caused the costs of scientific projects to skyrocket. While thrilling a larger audience with a vision of what trans-spectral astronomy beyond the

atmosphere can accomplish, he complained that servicing the Hubble Space Telescope from the space shuttle will cost almost as much as building a second telescope.

Among the grand questions, that of how cold it will be after a nuclear war remains alive, even though S. Fred Singer, a physicist at George Mason University, Virginia, heaped criticism on models predicting severe cold for their failure to take account of the properties in the infrared of the smoke and ice-clouds that would form after a nuclear exchange. He argued that there would be a greenhouse effect, as a result of water vapour carried into the stratosphere coupled with smoke lower in the atmosphere, huge enough to turn nuclear winter into nuclear summer.

But Stephen Schneider, of the National Center for Atmospheric Research in Boulder, Colorado, accused Singer of raising an "infrared herring". Schneider says Singer's physical issues do not change predictions that temperatures would drop after a nuclear exchange. Schneider says a low smoke cloud's infrared properties would not be significant, although the ice issue may have some relevance. But, in any case, said Schneider, even a relatively small drop in temperature could be disastrous if it produced a frost during the growing season. Schneider believes a more serious issue yet to be given the attention it deserves is the effect of fog on post-war temperature.

Pre-war as well as post-war problems were examined if not solved at AAAS. On arms control, Roger Fisher, a professor at the Harvard Law School, said at one point during five days of discussion of security issues that US negotiators may be good at presenting positions, but are less good at plotting strategy for making progress. One way to succeed is to describe new positions without commitment. In the same vein at a different meeting, Professor Thomas Schnelling, also of Harvard, had asked for a "moratorium on arms control negotiations" not only because of the damage they can do but because they divert attention from other beneficial courses of action.

Of necessity, there were not many jokes in these sessions, and even Academician R.Z. Sagdeev's joke was a black one. Seeking a simple model of the doctrine of mutually assured destruction, he told of the Russian couple who, when asked at their silver wedding how their marriage could have survived all the tribulations of married life, replied that they had filled their basement with dynamite, and that each had had a button.

John Maddox & Joseph Palca

Czechs agree with Poles over water pollution

London

THE Czechoslovak Minister for Forestry and Water Management, Franktisek Kalina, last week officially "deplored and apologized for" the effects of the oil spillage at the Kunice cement works last November, which polluted the Oder and caused considerable damage in the Polish province of Katowice. After a two-day meeting in Krakow with the Polish Environment Minister, Stefan Jarzebski, Kalina said that steps had been taken to prevent a similar spill in the future, that both sides had agreed to the expansion of joint monitoring of border and trans-boundary rivers and that the obligation to pay for cleaning up pollution would fall on the country from which the pollution had come.

Pollution of Czechoslovak origin is a long-standing grievance to the Poles. In the mid-1970s, when Polish censorship imposed strict controls on discussion in the media of industrial degradation of the environment, the ban did not apply to pollution originating from Czechoslovakia. In the case of the recent spill, there were two main causes of tension. One was disagreement between the two sides about the extent of the spill; the other was that the Czechoslovak authorities had failed to provide the Poles with adequate and timely warning of the spill.

Although, on 9 December, the Polish government press spokesman, Mr Jerzy Urban, dismissed western reports that the spill had caused a major "conflict" between the two socialist states, a few days later the Warsaw correspondent of Czechoslovak television, Ratislav Baja, noted Polish complaints that the Czechoslovak authorities had provided only "incomplete" and "distorted" data. Similarly, on 19 December, the Polish Deputy Premier, Jozef Koziol, answering questions from the Parliamentary Commission on Environmental Protection, described the information provided by the Czechs as "strongly understated".

The Czechoslovak answer was that the Poles had measured the spill wrongly, as what they had collected was not oil but an oil-water mix — and that some of the pollution was probably of Polish origin anyway. There has been no attempt by the Czechs to explain the delays in information and, ironically, on the eve of the Krakow meeting, the Warsaw daily *Zycie Warszawy* reported yet another Czech spill near the city of Ostrava, which although not large was once again disquieting, as the Polish environmental services first learned about such events "from chance observers".

Vera Rich

Towards truth in advertising in US university education

Washington

WILLIAM J. Bennett, Secretary of Education in the Reagan Administration for the past two years, is a name with which to inflame academics. Bennett's pitch is that the quality of higher education in the United States is deficient because academics neglect their primary responsibility (to teach) and that universities are bilking the US economy by increasing tuition fees unjustifiably. But he says that the *New York Times*, not he, was responsible for the word "greed" in the headline above a statement of his case last week.

Contrary to academic suspicion, Bennett does not have horns growing out of his head. His special unpopularity on the campuses of the United States may stem from his own academic background — Williams College, a PhD in philosophy at Austin, Texas, a law diploma at Harvard Law School and some years of teaching and administration at Boston University, President John Silber's centre of successfully self-proclaimed excellence; apostates are rarely loved. In 1981, Bennett was President Reagan's first appointment as chairman of the National Endowment for

the Humanities.

Bennett makes no bones about his objectives: "I'm deliberately setting out to improve the quality of higher education", but quality is carefully defined. Bennett says the essence of his complaint is that universities are failing to deliver what they promise. He says there is a huge gap between the rhetoric of university catalogues ("we collect them here") and the actual experiences of undergraduates. He cites as evidence the Boyers report from the Carnegie Commission for Higher Education last November, with its account of the incidence of student dropouts and of discontent among those who stay.

If colleges accurately advertised their wares, one gathers, the secretary would be content, knowing that institutions offering second-quality goods would then not find students. In the marketplace for higher education, Bennett is the campaigner for better advertising standards. "To thine own claims be true" is his motto.

But are not many institutions providing an excellent education? Bennett says his complaints apply to "most places", not to all. Faced with the example of one university given a glowing report card by an external invigilating group, Bennett quickly says "I'll take some credit for that". After two years of demanding systematic evaluation of universities, "I think we've won that argument".

On research, Bennett deplores the ease with which universities can upgrade their status, duplicating each other's efforts and increasing the demands on federal funds for research grants. "North Carolina used to have two universities, now it has seventeen", he says, offering as an illustration of the research contagion the way in which state system colleges give themselves airs above what used to be their station.

Bennett is offended by the "higher education lobby, which has done a very good job", and almost sympathetic to the US Congress, "which has to face it year after year". "If you want to use the word greedy", the lobbying on behalf of universities during the debate over last year's Tax Reform Act is where it applies, he says.

In the event, the private universities represented by the American Council on Education appear to have lost most of their claims for tax exemption for charitable gifts and the capital appreciation of property, while not all scholarships and fellowships will in future be tax-exempt.

In reality, Bennett says, it is "balderdash" for the universities to act poor-mouth when higher education has an aggregate income of more than \$100,000 million a year. His constant theme is that

there would be plenty of money for the universities if only it were wisely spent.

Bennett's draft budget for the financial year beginning next October would make the federal government, which even now contributes only 6–7.5 per cent of the total, an even smaller player. The administration hopes it will be allowed to save \$5,500 million from educational programmes, but the Secretary of Education seems less than confident that Congress will allow him to spend so little.

There will be a bitter argument over financial support for university students. Bennett makes no bones of his objectives; access for young people likely to benefit from higher education, but not at the expense of "taxpayers who never go to university". He declares that there cannot be



William J. Bennett — iconoclast or populist?

much wrong with a system in which 60 per cent of young people at least begin on higher education courses of some kind, but insists that, in equity, the burden of student support should be shifted from the present system of modest grants and interest-subsidized loans to one in which interest would not be subsidized but the rate of loan repayment determined by a graduate's salary.

Bennett will not be discomfited by the budget battle ahead. Down-to-earth though he may be, there is a touch of the messianic about him, as if he has a vision of how the universities of the United States could prosper in a free market for higher education. If unsurprised that most academics do not share his vision, he talks as if it is only a matter of time before they are converted. He has one powerful influence on his side — the knowledge that in the climate of disappointment that marks the beginning of the next presidential election campaign, the charge that the universities of the United States have shortchanged their clients may strike a popular chord. But Bennett has a soft spot for the problems of research in the humanities, including Greek and Latin, which may yet win him unexpected friends. **John Maddox**

New Indo – Soviet space agreement

Bangalore

As part of an agreement between the Soviet Union and India to improve telecommunications between the two countries, the Soviet posts and telecommunications ministry will give technical help in setting up an Intersputnik Earth station in India. India is also likely to lease transponders on a Soviet communications satellite, which will make possible an exchange of television and radio broadcasts.

During the visit of the Soviet leader, Mr Mikhail Gorbachev, to India at the end of last year, the Soviet delegation suggested the establishment of a space centre in India to train cosmonauts from developing countries.

India has also concluded an agreement with the Soviet agency Glavcosmos for the launching this year of India's three-axis-stabilized operational Earth observation spacecraft, IRS, from the Soviet Union. India is said to be the first ever customer to hire the launch service offered by Glavcosmos and this will be the first commercial launch of an Indian spacecraft from the Soviet Union. India's first satellite, Aryabhata, and the experimental Earth observation spacecraft in the Bhaskara series were put into space free of charge by the Soviet Intercosmos vehicle.

Radhakrishna Rao

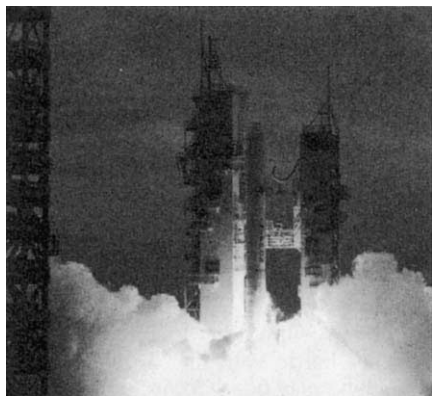
Japan's successful launch of first marine observation satellite

Tokyo

JAPAN's first marine observation satellite, MOS-1, was successfully launched from the National Space Development Agency's (NASDA) space centre in Tanegashima island off Kyushu on 19 February. The launch marks the end of an era of total dependence on US rocket technology and gives another boost to the morale of Japan's space engineers.

The satellite, named Momo (peach blossom), lifted off at 10.23 a.m. local time aboard a two-stage N-II rocket. After passing over Australia and Antarctica, the 740-kg satellite separated from the second stage over South America and entered a Sun-synchronous polar orbit at an altitude of just over 900 km about an hour after launch. Signals from the satellite were picked up immediately after separation at Kourou station of the European Space Agency in French Guiana.

Momo is equipped with three scanners: a multi-spectral electronic self-scanning radiometer with 50-m resolution in the visible and near-infrared to monitor water turbidity, red tides, waves and ice distribution over the sea, and vegetation, geology, water resources and snow cover over land; a visible and thermal infrared radiometer to scan snow, cloud, ice and water temperature distribution; and a microwave radiometer that will make cloud wa-



ter content, rainfall and snowfall measurements. The scanners lack the resolution of France's SPOT-1 satellite, which can pick up objects 10 m across when in panachromatic mode. Momo is the world's only marine observation satellite following the demise of the US's Seasat-1 shortly after launch in 1978. And its Sun-synchronous polar orbit allows coverage of almost the entire surface of the globe, bringing the satellite back to the same spot every 17 days. In addition to NASDA's Earth Observation Center north of Tokyo, ground stations in Spain, Norway, Thailand and the Antarctic will receive data directly from the satellite.

The N-II rocket that carried the satellite

aloft is based entirely on US rocket technology including the inertial guidance system which is a 'black box' taken from the US Delta rocket. But this is the last N-II launch and in future NASDA will use rockets largely or entirely developed in Japan. The H-I rocket, which was successfully launched last summer, has only US components in the first stage; the H-II, due to be completed in 1992, will be entirely made in Japan.

NASDA executive vice-president S. Sonoyama is noncommittal on whether Japan will enter the commercial space race with the H-II. But other officials know they must reduce launch costs for geostationary satellites from the present level of Y30 million (\$200,000) per kg for the H-I to Y8 million per kg to be commercially viable.

The H-II will have a payload capacity of only 2,000 kg for geostationary satellites, half the capacity of Ariane-5. But Sonoyama believes conventional geostationary satellites will not rise above 2,000 kg because of difficulties of launching heavy satellites; instead he sees the future in multipurpose space platforms.

The Ministry of International Trade and Industry, the Institute of Space and Astronautical Sciences affiliated with the Ministry of Education, Science and Culture, and NASDA are jointly developing an experimental space platform that will be launched by 1993 at the earliest by an H-II rocket and brought back by the US space shuttle. Although current specifications for the H-II limit it to satellites up to 3.7 m in diameter, plans are being made to increase this to 4.6 m for space shuttle compatibility. And the H-II could launch large payloads (up to 4,000 kg) by augmenting the second-stage engine and adding boosters. Thus the next generation of satellites does seem within Japan's future launch capabilities.

The only remaining barrier to Japan's entry to the commercial market is the fishermen of Tanegashima, who restrict NASDA to two 45-day launch seasons in summer and winter, claiming that falling rocket boosters damage their fishing grounds. But even this may not be a problem. Masashi Motchizuki, director of the H-II launch vehicle group, confidently predicts that 10 launches could be made during the summer season alone.

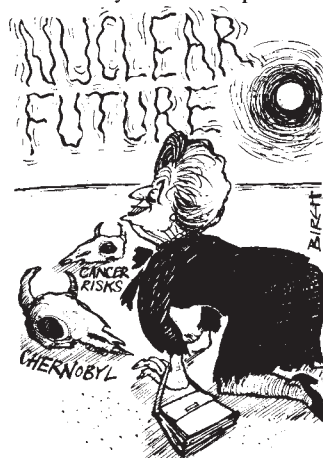
Since Japan launched its first satellite in 1970, there has been only one failure out of 35 satellites, an experimental communications satellite launched in 1979. In this age of space disasters and soaring insurance costs for satellites, reliability will probably be Japan's strongest selling point.

David Swinbanks

Safety fears in nuclear debate

London

THE safety risks of nuclear energy have still not been properly assessed and there is no proper emergency procedure to cope adequately with another Chernobyl. These fears surfaced in public debate last week as British parliamentarians from all political parties expressed the growing disquiet among some members of the public about the safety of nuclear power.



At the beginning of the week, the House of Commons debated the findings of the 109-chapter report, prepared by Sir Frank Layfield, on the proposed pressurized-water reactor (PWR) to be built by the Central Electricity Generating Board (CEGB) on the Suffolk coast. Despite the opposition, there is little doubt that the government will approve the reactor, estimated to cost about £1,300 million.

The leader of the Liberal Party, Mr David Steel, was typical of those who consider the conclusions of the Layfield report on safety to be inadequate. In a letter to Mr Peter Walker, the Energy Secretary, Steel demanded that the risks of civil nuclear power be assessed and made public: "It [the report] appears to suggest that Sizewell B will be acceptably safe only if the economic benefits are large enough to justify the scale of the risk involved. Or, put even more directly, that there are safety risks involved but that these will be acceptable if the economic benefits are big enough."

The public debate on Sizewell B resurrected the fears about Chernobyl. Although the National Radiological Protection Board (NRPB) tried to inform all the local authorities seeking guidance in the immediate wake of the disaster, the "arrangements" to cater for such an emergency proved inadequate.

As a consequence, NRPB has advised the Department of the Environment, which is drawing up the new procedures, to create a national emergency telecommunications network.

Bill Johnstone

India establishes countrywide satellite computerized database

New Delhi

A NATIONWIDE satellite-based computerized information network is being set up by the Indian government to provide its planners with instant access to information needed for decision-making and monitoring of development programmes.

By the end of this year, government offices in each of the 430 districts, the 22 state capitals and major cities will be hooked into a vast information network that will move data from virtually any part of the country to the offices of ministers and their secretaries in central and state governments.

The man behind the \$70 million project is Dr N. Seshagiri, director of the National Information Centre (NIC) created by Mrs Indira Gandhi in 1977. NIC now services 62 central government departments in New Delhi through a cable and VHF network linking computer terminals in individual departments with NIC's 170/720 computer, the biggest in India, which stores some 200 databases covering crucial sectors of the economy.

NIC's computerized management information has already proved valuable. For example, the performance of power plants is monitored daily, trade figures are updated every day and performance indices of all public-sector undertakings are regularly tracked. The 1987 budget, to be presented this month by the Prime Minister, Mr Rajiv Gandhi, will also be based on data drawn from NIC's computer. According to Seshagiri, NIC's services are now being extended to cover the whole of India with the help of the satellite.

The new project will ensure the creation of district databases on health, agriculture, rural development, employment and

other relevant social and economic matters; they will be stored in locally made microcomputers in the offices of district collectors. Four large mainframe computers (NEC-1000) imported from Japan have already been installed in Delhi, Pune, Bhubaneswar and Hyderabad to store databases from the regions.

Within the next two months, superminicomputers will be installed in the offices of the chief secretaries in all the state capitals. These computers are being made by the state-owned Electronic Corporation of India under licence from the Control Data Corporation of the United States.

Except for the imported master Earth Station at New Delhi, all other computers in the network will be linked to the satellite by micro Earth stations and two-way roof-top antennas. The first batch of 120 micro Earth station antenna systems are being imported from the Equatorial Corporation in the United States. The rest will be supplied by Indian Telephone Industries in Bangalore, which has purchased the technology from the US company.

According to Seshagiri, the network will become operational in May, using a leased Intelsat transponder over the Indian Ocean. But NIC hopes to get a loan from the World Bank to acquire and launch its own satellite, Nicsat, in collaboration with the Indian Space Department, by 1990. Nicsat will operate in the Ka band (20 GHz), making possible the use of still smaller (50-cm diameter) dish antennas costing less than \$2,000 each.

Seshagiri believes India is the first country to establish a satellite-based information network dedicated to government decision-making and to toning up the sag-

Schools wooed for science teaching

London

THE promotion of science in British schools was given a boost last week with the launch of two programmes to attract unconfident teachers to science and to encourage them to include it in their curriculum.

The first is a comprehensive basic science teaching kit in the form of teachers' notes and colourful learning cards for children. The kit, prepared by the British Gas Corporation, and designed for young children, will outline the basic concepts of air, flight, forces, heat, water and 'ourselves'.

Another project is designed to promote education in science using information stored on a computer database. The project, which will be funded by the Department of Trade and Industry until April next year, allows teachers with microcomputers and modems to 'dial' into the system to obtain teaching aids.

The system, called the National Educational Resources Information Service, will have about 500,000 'resources' listed on the database by 1988. The data will be housed in the computer of the Open University.

Bill Johnstone

ging administrative machinery. "Some 10,000 government-aided projects are being implemented in various districts, but central ministers have no fast means of monitoring the time and cost overruns", says Seshagiri. He says the cost of the network will be recovered several times over by better management of multimillion dollar development projects that now languish because of delays in decision-making.

States that were initially hesitant to share information with the centre have joined the network after an assurance from NIC that the autonomy of the federal structure will be respected. No data relating to police or security matters or that violates the privacy of individuals or organizations will be collected. Seshagiri says 85 per cent of information in databases will be open to officials; the remaining 15 per cent will be secure and available only to specially authorized personnel. About 1,300 government officers have been trained in database creation and NIC hopes to train several hundred more soon.

At the same time, the state-owned Computer Maintenance Corporation of India is setting up a network called Indonet for public use. It provides computer booths throughout the country that can be used by, among others, small businesses unable to afford their own computers. The fourth-generation IBM computers at Bombay, Hyderabad, Delhi, Calcutta and Madras that form Indonet will eventually be linked via satellite and expanded to cover 35 cities.

K.S. Jayaraman

Teacher union issues AIDS guidelines

London

THE National Union of Teachers (NUT), the negotiating body for most of Britain's schoolteachers, has published guidelines on AIDS (acquired immune deficiency syndrome) to "reassure members and dispel myths about the spread of the disease".

According to the NUT deputy general secretary Doug McAvoy: "It is vital that teachers know the facts. The NUT's notes on AIDS should reassure members that they need not worry about contracting AIDS in their contact with pupils." Teachers should be entrusted with confidential information on children's health, and confidentiality must be maintained to protect children from victimization as a result of unfounded fears of infection.

The guidance notes, to be distributed to

the NUT's 650 local associations and branches, focus on homosexual and bisexual men, drug addicts who share needles, and haemophiliacs as the 'risk groups'.

Evidence shows, says the union, that children who carry human immunodeficiency virus pose no risk to other children or adults in a normal school environment and should attend school normally.

The union says: "None of the identified cases of HIV infection in this country or the United States is known to have been transmitted in a school setting, nor has the virus been spread through casual personal contact. There is no reason for teachers to be concerned about handling objects used by an infected person and no need for separate toilets or crockery and cutlery".

Bill Johnstone

Antimatter underestimated

SIR—The first published account of success in immobilizing antiprotons in an electromagnetic trap has been the subject of a recent leading article in *News and Views* (*Nature* **324**, 299; 1986). Although this article gives a good overview of the scientific meaning of this historical achievement, it gives in our view a rather misleading perspective on its likely military significance.

It is true that for some applications of antimatter, such as antiballistic missile or spacecraft propulsion, relatively large amounts of antiprotons are required. For most other cases, the useful amounts are usually much smaller. For example, we have calculated that less than a microgram of antiprotons is sufficient to trigger a thermonuclear explosion or pump a powerful X-ray laser^{1,2}.

But antimatter is not only the most portable of all high explosives, it is also the only feasible portable source of muons. In every antiproton annihilation, on average three muons are produced. These could be used to induce muon-catalysed fusion reactions in a deuterium-tritium mixture, an attractive solution for a low-weight space nuclear reactor that could be operated in a continuous or pulsed mode.

Furthermore, by collecting and cooling the muons (an easy task compared with that of cooling antiprotons) a very intense beam could be formed and sent into the atmosphere to guide, over a range of more than 10 km, a series of powerful electron or proton beam pulses towards a target. More simply, stopping the muons in a suitable material would generate an extraordinarily effective X-ray lasing medium, for the two-microsecond lifetime of the muonic atoms.

In outer space, a very low intensity burst of antiprotons would be most suitable for active warhead/decoy discrimination. In this and the two previous examples, the amount of antiprotons needed is of the order of *nanograms* per engagement. Conservative estimates of the technical problems involved in producing and manipulating *microgram* amounts of antimatter per day show that known technology is only a couple of orders of magnitudes away from meeting the challenge^{2,4}.

We are very much concerned by the implications for nuclear weapons proliferation of the indisputable scientific feasibility of several antimatter weapon concepts.

To develop such weapons would add considerable impetus to the current arms race. We call for an immediate ban on all antimatter-related research, especially as this work is fundamental to many third-generation nuclear weapon systems.

ANDRÉ GSPONER
JEAN-PIERRE HURNI

*Independent Scientific Research Institute,
15, Charles-Galland,
Ch-1206 Geneva,
Switzerland*

SDI boycott

SIR—Your leading article (*Nature* **323**, 746; 1986) questioning the wisdom of a boycott of the Strategic Defense Initiative (SDI) misses the point. Scientists who oppose SDI funding do not accept the premise that knowledge is worth gaining for its own sake, irrespective of the ultimate social consequences. We can wait to obtain some knowledge and its gains.

Of course academic researchers should decide for themselves what research they should undertake, but they should also strive to ensure that the knowledge they gain is not applied to ends that are not in the interests of humanity, questions of nationalism aside. Refusing SDI funding is one way of expressing opposition to the political control of science.

We in New Zealand admire the refusal mentality and although there is no prospect of our being offered SDI money, scientists have played an important role in bolstering the government's policy of refusing port visits by nuclear weapon-capable vessels, of the Royal Navy and the United States Navy in particular.

PETER WILLS

*Department of Physics,
University of Auckland,
Private Bag,
Auckland,
New Zealand*

In the greenhouse

SIR—T.R. Vidyasagar (*Nature* **323**, 390; 1986) suggests a switch to a vegetarian diet and reforestation of the 40 per cent of present agricultural land so freed as a way to decrease atmospheric CO₂. This short-term measure could be supplemented by fertilizing the world's forest and pasture soils, where applicable, with nutrients from city drains, combined with pelleted spores and seeds of soil-enriching plants. Only a light dressing of suitably prepared and selected material would be required to fortify their nutrient cycles and make them more productive for many years. Many parts of the oceans would respond to similar treatment by producing more fish, seaweed, corals and carbonaceous

sediments, the latter being the more permanent sink for CO₂.

Perhaps floating lagoons made to transport sewerage could be towed like icebergs (used for soft water irrigation), to chosen sites, where the surface waters would be fed by trickle and slow release methods. Natural upwelling of nutrient-rich bottom waters of the ocean greatly increase their production of living organisms, (J.H. Ryther *Science* **166**, 72–76; 1969) and artificial methods could be developed to extend this process.

I look forward to a time when it is possible by careful balancing of the nutrition of pH of the biosphere in these and other ways to treble its content of unoxidized C, as has been accomplished on farms and gardens around the world in the course of a few decades.

R. PARR

*Arrington Bridge,
Arrington, Royston,
Hertfordshire SG8 0AE, UK*

First among equals

SIR—in a recent leading article (*Nature* **324**, 509; 1986), you ask “Should editors now seek to protect authors from themselves by banning the use of the word ‘first’?”. My reply is: would you care to join the *Proceedings*, which has had such a policy for some years?

I do not know exactly when the *Proceedings* first published priority statements. There is a 1977 memorandum in the files from Bernard K. Forscher, then managing editor, to Robert L. Sinsheimer, then chairman of the editorial board, suggesting such a policy (which may simply be the earliest available written expression of a previously unwritten policy) and a letter from Sinsheimer concurring. Since 1983 at least, the Information for Contributors has stated “The *Proceedings* does not print priority statements”. Thus, one of the duties of our editorial staff is to remove such statements. Most authors accept this change with equanimity; for those who do not, we permit the qualified statement “So far as we know, this is the first...”.

A policy that is much more difficult to implement is our proscription of statements of novelty. I admit that this has been unevenly enforced, although in the past year we have become more strict. Novel fragments, novel genes, novel hypothesis — the term is usually a cliché occasioned by the author's desire for an adjective, as our stated policy is to publish reports of exceptional importance and novelty. Again, would you care to join us?

FRANCES R. ZWANZIG
(Managing Editor)

*Proceedings Office,
National Academy of Sciences,
2101 Constitution Avenue,
Washington, DC 20418, USA*

1. Gsponer, A. & Humi, J.-P. *La Recherche*, **17**, 1440–1443 (1986); *The Physics of Antimatter Induced Fusion and Thermonuclear Explosions*, in *Proc. 4th Intl. Conf. Emerging Nucl. Energy Syst.*, Madrid, 30 June–4 July 1986 (in the press).
2. Gsponer, A. & Hurni, J.-P. *Antimatter Induced Fusion and Thermonuclear Explosions*, report ISRI-85-05 (August 1985); preprint ISRI-86-06 *Atomkernenergie/Kerntechnik*.
3. Augenstein, B.W. *Concepts, Problems and Opportunities for use of Annihilation Energy* RAND Note N-2302-AF/RC (June 1985).
4. Walgate, R. *Nature* **322**, 678 (1986).

Molecular genetics of the mind

The use of DNA markers has shown that manic-depressive illness can be caused by a single gene. These markers will not be useful for screening but may lead to fundamental insight into the disease.

THERE has for some time been good evidence that individuals in some families may be genetically predisposed to a spectrum of mental diseases of which the most distinctive is manic-depressive psychosis. About two years ago, Daniela Gerhard and colleagues published an abstract suggesting that manic-depressive disease among the Old Order Amish of Pennsylvania might be traceable to a single gene on the short arm of chromosome 11 (Gerhard *et al. Am. J. hum. Genet.* **36**, 35; 1984). Evidence in support of this suggestion is reported by Egeland *et al.* on page 785 of this issue. On pages 805 and 806 the assiduous reader will find persuasive evidence from two other research groups for an inherited predisposition to manic-depressive disease that is not linked to the short arm of chromosome 11. None of these groups has actually identified the predisposing gene.

It is not in principle surprising that a defect in a single gene may cause a complex and variable disorder of human behaviour, or that the predisposition is apparently attributable to different genes in different families. It is, however, remarkable that it has been possible to establish the linkage to chromosome 11 in one of them, for reasons connected partly with the nature of the disease and partly with the more general obstacles presented by the human species to the study of genetics.

For much the same reasons, there are strict limits on the practical uses to which the new genetic markers for manic depression could be put. Manic-depressive illness (also known as bipolar affective disorder) is relatively common, affecting an estimated 0.5 to 1.0 per cent of people. Although manic-depressives characteristically undergo alternating cycles of wild excesses and profound depression, the age of onset, frequency and severity of the mood changes vary considerably, and some individuals in affected families manifest only depression. Moreover, while the effect of the gene is dominant — that is, an individual carrying only one copy can be affected — it has incomplete penetrance, which means that some fraction of those carrying it will never develop the disorder.

Genetic analysis of such a disease faces obvious difficulties. First, the diagnosis of the disorder necessarily requires subjective assessment; second, if more than one gene is involved, genetic evidence from different families cannot necessarily be pooled; and third, incomplete penetrance

may make it very difficult to establish a clear link between the inheritance of a given gene and the inheritance of the disease. In any case, the unequivocal establishment of such linkages requires large and generally rather inbred families, which are rare in Europe and North America.

These last conditions are however met by the Old Order Amish whose religious convictions moreover forbid alcohol and drugs, thus eliminating one serious obstacle to the accurate ascertainment of psychiatric disease. What Egeland *et al.* have been able to establish, in brief, is that in one large Amish family, two genes known to lie close together on chromosome 11 are frequently inherited together with affective disorder, which in their study was rather stringently defined and independently diagnosed by five independent assessors.

They were able to establish this linkage because the cloned genes they used as markers lie adjacent to highly variable regions of DNA, making it extremely unlikely that a given individual will inherit identical variants of this region from both parents. By tracing the inheritance of different variants of the two marker genes, and of affective disorder, through the affected family, Egeland and colleagues were able to show that a gene predisposing to manic depression lies close to the two marker genes on chromosome 11.

What are the implications of this linkage, especially in the light of the evidence of the other two groups, Hodgkinson *et al.* and Detera-Wadleigh *et al.*, that in three Icelandic families and three North American families respectively the chromosome 11 markers used by Egeland and collaborators are not associated with affective disorder? First, although mistakes in diagnosis could more easily lead to a false negative conclusion than a false positive, all three research groups used the same standard diagnostic criteria and the negative evidence is quite convincing. This means there are at least two different genes predisposing to affective disorder.

In fact there are probably at least three: a large Israeli survey using traditional genetic markers provides strong support for earlier suggestions that some cases may be due to a gene on the X chromosome (M. Baron *et al. Nature*, in the press). Although it seems quite likely that most cases of manic-depressive disorder arise from a genetic predisposition, it is

clearly impossible to tell at this stage how many will be linked to chromosome 11. Genetic heterogeneity alone will thus limit the usefulness of the chromosome-11 markers in identifying individuals at risk in the general population. But in any case the use of linked markers for this purpose is generally impossible without unusually extensive family data, as the particular variant of each marker that is linked to the disease will differ from one family to another. Moreover because of the incomplete penetrance of the gene (Egeland *et al.* estimate maximum penetrance at 63 per cent), there is no guarantee that an individual carrying the predisposing gene will ever develop the disorder.

Finally, what kind of single gene might produce the spectrum of complex disorders that seem to arise in affected families? Egeland *et al.* point out the highly suggestive recent finding that their two marker genes on chromosome 11 are closely linked to the gene encoding tyrosine hydroxylase, which is the principal enzyme in the synthesis of the catecholamines. The catecholamines comprise a major class of neurotransmitters, disturbances in whose function are implicated in a wide range of mental disorders.

It is not difficult to imagine ways in which abnormalities in the regulation of either the enzyme or the gene encoding it might lead to imbalances in neural activity that could, by inducing delayed compensatory changes in other systems, result eventually in periodic oscillations expressed as manic-depressive cycles. Nor is it hard to see how such regulatory defects might occur either through mutations in the tyrosine hydroxylase gene itself or in genes encoding (perhaps) tissue specific regulators of the gene or the enzyme.

At this stage, so little is known about either the gene or the development and regulation of neural-signalling systems as to leave the imagination dangerously unfettered. However, Moss *et al. (Nucleic Acids Res.* **14**, 9927; 1986) have recently reported variable regions of DNA in the vicinity of the tyrosine hydroxylase gene. Hodgkinson *et al.* used the variation to look for linkage with manic-depression in the Icelandic families and drew a blank. It is a very safe bet that a similar investigation will soon be reported in the Amish; and that the genomic organization and regulation of the tyrosine hydroxylase gene will not be left to the imagination for long.

Miranda Robertson

Solid-state physics

Liquid-nitrogen-temperature superconductors arrive

John Maddox

EVEN before last week's issue of *Nature* had been mailed to subscribers, last week's record high temperature for the transition to superconductivity, given by Strongin, Welch and Davenport as roughly 70 K (*Nature* 325, 664; 1987), had been broken. On Monday, the *New York Times* announced that C.W. Chu and his associates at the University of Houston, last week's record-holders, had succeeded in preparing a ternary oxide that is superconducting at 98 K, well above the boiling point of liquid nitrogen at 77 K. One excited physicist wondered whether the US Department of Energy had not been overhasty in agreeing to build a proton-proton collider using helium-cooled magnets.

These developments are every bit as important as Strongin, Welch and Davenport explained last week, when they also warned that it will be some time before these materials find their way into quite small devices, let alone into the power-carrying circuits of the big machines. During the past few months in which groups in the United States and Japan have been racing toward the goal of a liquid-nitrogen superconductor, they have had to make the best they can of specimens a few millimetres across which vary markedly in their properties from one to another.

Chu's technique (see Chu, C.W. *et al. Phys. Rev. Lett.* 58, 405; 1987) requires that a mixture of the oxides of lanthanum, of calcium, barium or strontium, and of copper should be heated for hours on end at 900 °C and that the solid mass produced should be ground and sintered at 1,200 °C for another spell. There is no assurance that the resulting specimens are uniform solids, let alone regular crystals. Indeed, it has been possible to show that many of them are mixtures of two phases, each with a different crystal structure. Most specimens appear to be replete with empty voids, hardly suggestive of the purity and order that might be thought necessary for a superconductor. One of the goals ahead is the preparation of these materials in something like a well-ordered form.

What are they? And by what stretch of the imagination can an oxide be a superconductor? In reality, the ternary oxides (or sulphides, for that matter) are more often semiconductors, which is credible enough. But one of the surprises of the past five years or so is that small changes of physical conditions or of chemical composition can profoundly change the nature of a material. Some may be ferromagnets

at low temperatures and semiconductors above some critical point. Ferroelectric characteristics are common.

The oxides in the forefront of people's minds in the past few months are materials that should ideally form crystals in which the two different kinds of cation (copper on the one hand and lanthanum and barium/strontium/calcium on the other) form planar sheets in each of which the atoms are linked together by oxygen atoms and which are then stacked on each other, first one kind and then the other.

Put differently, each copper ion is surrounded by a distorted octahedron of oxygen anions, giving tight binding within each copper plane but much looser binding in the direction perpendicular to the planes. The materials on which most interest has centred in recent weeks may be thought of as derived from LaCuO_3 or La_2CuO_4 by the stoichiometric substitution of group II elements such as calcium for lanthanum atoms. But neither of the starting materials is a superconductor.

What seems to happen is that, especially in the short bonds between copper and oxygen atoms within the planes, bonding electrons acquire an unusual degree of delocalization, and hence mobility. On the face of things, the situation is a little like that in graphite, in which again there is a layered structure (based on a hexagonal network of carbon atoms) loaded with potentially mobile electrons — and it is no accident that there have been several attempts in the past few years to detect superconductivity in graphite at low temperatures and high pressures, all of them unsuccessful.

Part of the reason for this is that graphite is strictly a semiconductor, but one without a finite band-gap in the usual meaning of the term; instead, the density of electronic states as a function of energy tapers to a vanishingly small waist at the Fermi level, the energy of the most energetic occupied state when all the available electrons are fitted in most economically. The vanishingly small density of states at the Fermi level militates against superconductivity, dependent as that is on the interaction between mobile electrons and lattice vibrations. The electrons that matter will be those lying near the top of the occupied band, which can be excited into a conduction state by amounts of energy as small as those of lattice vibrational quanta (phonons); in graphite, the number of electrons available for conduc-

tion will be vanishingly small. Turning the argument around, the trick in making an oxide superconductor is to arrange that the density of electronic states as a function of energy is thick-waisted at the Fermi level.

The density of states at the Fermi level is not the only criterion that must be satisfied for superconductivity at high temperature to occur. It is also necessary that the electron coupling should be close, which is assured in these oxides by the proximity of neighbouring copper ions and by their capacity to exist in doubly or trebly charged form, which allows the exchange interactions to be strong. What remains to be puzzled out is why the coupling between the electron states and the lattice vibrations should be as strong as it appears to be; for that purpose, accurate crystal structures would be a boon, but they may not be easy to come by with materials made by sintering.

Several influences seem to matter in the design of the new materials. The substitution of a group II atom for lanthanum appears to be largely a way of influencing the electronic balance in the copper-carrying layers on either side, perhaps changing the electron affinity of the copper ions. In a sense, the group II elements among the lanthanums may be likened to the influence of the voltage on the grid of an old-fashioned triode valve.

Another way of changing the density of electronic states in the copper planes appears to be a stoichiometric deficiency of oxygen atoms, which explains why the properties of the oxide superconductors so far prepared appears to be sensitively influenced by the atmosphere in which they are prepared. In some cases, a reducing atmosphere will yield a superconductor whereas an oxidizing atmosphere will not; in others, the reverse is true.

So far, it seems, these empirical truths have been learned the hard way. People have prepared samples with very different compositions and have plotted selected measured quantities on diagrams in the hope that contour lines will lead them to an optimum composition. Now it may be possible to provide some explanation of what has happened. That is one problem that cries out to be tackled.

Another is that of controlling the composition of these materials more accurately than is possible by sintering at high temperatures. For even if, for the time being, oxide superconductors will be used only in small electronic devices (will the superconducting magnetometers called SQUIDS be the first outlet for them?) the sheer mechanics of production will require uniform and reproducible raw material. That is another direction in which there will have to be progress if the potential of the new development is to be realized. □

John Maddox is Editor of *Nature*.

Muscle regulation

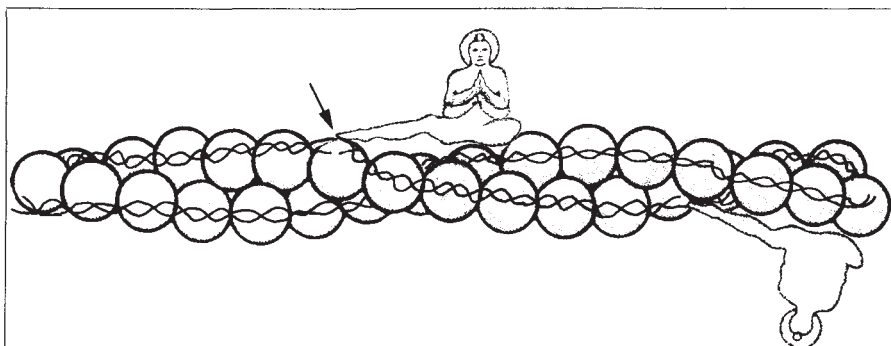
Movement on the *Aufbaubahn*

C. E. Schutt

IT is possible that muscle evolved by near avoidance of crystallization. Perhaps, eons ago, our first motile ancestors simply salted away under some condition or other, only to be revived by the timely arrival of ATP. Indeed, the beautiful lattices in insect flight muscle reveal how closely nature, in toying with the idea of using the drive towards symmetry as a force, came to going over the brink. Electron microscopy indicates that, in the rigor state, with no available free energy, each myosin head apparently rests in peace with an actin mate on the thin filament. Thus, force develops as these proteins constantly fall towards an equilibrium association, only to be brought back again by an influx of energy-rich ATP. This cycle is regulated in vertebrate muscle by the calcium-sensitive troponin/tropomyosin system of the thin filament which teases the myosin molecules in their relentless attempts to join with actin and release the products of ATP hydrolysis¹. Two recent crystallographic results^{2,3}, one of which is reported by White *et al.* on page 826 of this issue³, and some earlier biochemical studies^{4,5}, point to a new way of thinking about cooperative molecular movements associated with this process (see figure).

A curious aspect of tropomyosin structure, first revealed by pattern analysis of its newly determined sequence, is a 14-fold periodicity in the position of charged and nonpolar residues on the surface of a two-stranded α -helical coiled-coil^{6,7}. Because seven sites were expected on the basis of its alignment with actin monomers along the thin filament, these observations naturally supported the prevailing two-state on/off model for regulation, where the long tropomyosin molecule shifts *en bloc* to a new position on the actin helix when troponin binds calcium⁸. In this view, myosin heads either had a shot at an actin molecule or were blocked. Clearly, a more chemical understanding of this 'steric-blocking' mechanism would require a nearly impossible feat of crystallization to capture all the molecules *in flagrante delicto*. (The 'dream' crystal would have fourteen helically arranged actin monomers, bound to a tropomyosin/troponin complex on each side, with a complement of myosin heads thrown in.) Nevertheless it is remarkable that nearly all the principal actors in the contractile apparatus have been separately crystallized: actin^{9,10}; tropomyosin²; troponin-C^{11,12}; myosin heads (S1)¹³; and, now, fragments of troponin-T complexed with tropomyosin³.

It might someday be feasible to construct from crystal structures of these proteins a view at atomic resolution of the force-generating event. This is an example of the *Aufbau* problem of structural biology, namely how to build up pictures of higher-order structures from crystallographic descriptions of their components. The case of sickle-cell haemoglobin fibres illustrates how various experimental approaches can be used to work up from a monomer crystal structure to a model of the higher-order structure. In fact, a key



Troponin/tropomyosin teasing myosin with a flick of the leg. Circles represent the monomers of the helical actin filament. There are 14 monomers per helical repeat and one molecule of the troponin complex, represented by the buddha, binds at each cross-over point. Tropomyosin is shown superimposed on the actin filament; each molecule ends in a forked lap joint (arrow). On calcium binding, troponin-C, the torso of the buddha, untwists (see my previous News and Views article¹⁷ for details of this mechanism). Inhibition of the complex by troponin-I (the buddha's knee) is thus removed. Troponin-T (the opposite leg) is bound to the lap joint of tropomyosin at the amino-terminal end (see text for details of the regulation mechanism).

aspect in understanding that puzzle was the demonstration that sickle-cell haemoglobin monomer/monomer contacts seen in protofilaments were the same as those in the crystals¹⁴. Similarly, it might be that one of the two crystal forms of monomeric actin, DNase-actin⁹ or profilin-actin¹⁰, will show actin/actin contacts like those in thin filaments (but necessarily twisted axially to fit space-group symmetry). This would be a clear case of 'speeding along the *Aufbaubahn*'.

Likewise, tropomyosin crystals, those extraordinary watery meshworks of criss-crossing coiled coils, are held together by head-to-tail intermolecular contacts, lap joints, that must be very similar to those used in muscle⁶. In a tour-de-force (no pun intended) of isomorphous replacement, White *et al.*³ were able to exploit the mostly empty (except for water) lattices to soak in large, well-characterized, proteolytic fragments of troponin-T, the tropomyosin-binding subunit of troponin. A series of difference Fourier maps from these soaked crystals clearly shows the troponin-T amino terminus binding to the tropomyosin/tropomyosin lap joint (see

figure). From this result, the authors conclude that troponin-T strengthens the lap joint, thereby enhancing cooperativity among the set of linked tropomyosins wrapping slowly around the actin helix. The globular carboxy-terminal head portion of troponin-T pins tropomyosin to the edge of the actin groove at low calcium levels, preventing it from making regular contacts via the seven strong tropomyosin actin-binding sites³. These sites become occupied only at high calcium levels as successive myosin heads attach, inducing the fully 'potentiated' state⁵. (White *et al.* attach no particular structural significance to the second set of tropomyosin actin-binding sites found in the linear sequence.) Thus, the state of any particular stretch of regulated thin filament, in terms of sensed calcium levels, actin

conformation or numbers of bound myosin heads, is telegraphed all the way down the line, consistent with recent views⁴ of cooperative regulation, where single actin/myosin binding events are no longer equivalent and independent.

The tropomyosin crystallographers² would like to go a step further by carrying over to muscle not only the static picture afforded by the crystals but also important details of the molecular motions seen there. The rapid and non-isotropic fall-off of intensity as a function of resolution in the diffraction pattern indicates considerable movement of the coiled-coil about its mean position, the excursion being smaller at those points where the molecules are constrained by contacts with their neighbours. How would their picture of a highly flexible tropomyosin molecule fit with the notion of cooperative regulation? The largest motions are thought to occur in the unpinned state because, by analogy with tropomyosin crystals, a smaller number of constraints apply there. Thus, when the troponin switch is thrown, tropomyosin enters a continuum, riding rein on surging actin molecules at full gallop as myosin

heads struggle to mount up. As more and more myosins grab hold, a regular series of actin/tropomyosin contacts is made, establishing the fully on state. This image accords well with the dynamic movements seen in fluorescently labelled actin filaments¹⁵ and with recent time-resolved X-ray diffraction experiments on intact muscle fibres¹⁶. □

1. Taylor, E. W. *CRC Crit. Rev. Biochem.* **6**, 103–164 (1979).
2. Phillips, G. N. Jr et al. *J. molec. Biol.* **192**, 111–131 (1986).
3. White, S. P. et al. *Nature* **325**, 826–828 (1987).
4. Hill, T. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 60–64 (1983).
5. Bremel, R. D. & Weber, A. *Nature new Biol.* **238**, 97–101 (1972).

6. Parry, D. A. D. *J. molec. Biol.* **98** 519–535 (1975).
7. McLachlan, A. D. & Stewart, M. J. *J. molec. Biol.* **103**, 271–298 (1976).
8. Huxley, H. E. *Cold Spring Harb. Symp. quant. Biol.* **37**, 361–376 (1972).
9. Kabsch, W. et al. *EMBO J.* **4**, 2113–2118 (1985).
10. Schutt, C. E. et al. in *Actin: its Structure & Functions* (eds Yanagida, T. & Oosawa, F.) (in the press).
11. Herzberg, O. & James, M. N. G. *Nature* **313**, 653–656 (1985).
12. Sundaralingam, M. et al. *Science* **227**, 945–947 (1985).
13. Rayment, I. & Winkelman, D. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4378–4380 (1984).
14. Dickerson, R. E. & Geis, I. *Haemoglobin: Structure, Function, Evolution and Pathology* (Benjamin, London, 1983).
15. Yanagida, T. *Nature* **307**, 58–60 (1984).
16. Kress, M. et al. *J. molec. Biol.* **188**, 325–342 (1986).
17. Schutt, C. E. *Nature* **315**, 15 (1985).

C. E. Schutt is in the Department of Chemistry at Princeton University, Princeton, New Jersey 08544, USA.

Conservation genetics

The future of the giant panda

Stephen J. O'Brien and John A. Knight

THE giant panda has captured media attention and the fancy of westerners since its description by Pere Armand David in 1869. In recent decades breeding pairs have been presented by the Chinese government to visiting foreign dignitaries as tokens of friendship, but only 14 giant pandas reside outside China. Despite the considerable efforts of the Chinese people and government to conserve pandas in the wild, the population continues to decline and at the present rate the giant panda will become extinct perhaps as early as the twenty-first century. Last autumn Prince Philip (president of the World-Wide Fund for Nature, WWF) visited the Wolong Natural Reserve and its Conservation and Research Centre for the giant panda in the Sichuan province on the eastern edge of the Tibetan plateau in mainland China. His visit prompted a meeting at the reserve of Chinese and foreign scientists from several disciplines with one interest in common: the biology of the giant panda and its neighbours. Based on accumulated data, a draft of a plan for preservation of the species was prepared by a joint WWF–Chinese Ministry of Forestry team headed by John MacKinnon of WWF. We report here what the Prince learned from his meeting with the group.

Panda fossils have been uncovered in many locations in mainland China and South-East Asia. As recently as 2,000 years ago pandas inhabited five Chinese provinces (Hehan, Hubei, Hunan, Guizhou and Yunnan) where they do not occur today. The remaining pandas exist in six small pockets of mountainous land (about 30,000 km²), mainly in the Sichuan province. Pandas subsist on a nearly exclusive diet of bamboo shoots, stems and leaves and live in 2,000–3,500-m-high dense alpine bamboo forests. Twenty-one species of bamboo eaten by giant pandas in China have been identified. MacKin-

non and Robert de Wulf (State Univ. Ghent) recognize suitable panda habitat as bamboo forest (higher than 2,000 m) with greater than 70 per cent forest canopy, possessing at least one species of preferred bamboo. In the Wolong Natural Reserve the two preferred species of bamboo are *Sinarundinaria fangiana* (recently renamed *Bashania fangiana*), arrow bamboo which grows at higher elevations of 2,600–3,200 m, and *Fargesia spathacea*, umbrella bamboo at 1,500–2,600 m. On these criteria only 20 per cent of the present range of the panda (6,000 km²) remains as suitable habitat.

Between 1974 and 1977 about 3,000 Chinese scientists conducted a census of the remaining wild pandas in China, based on a transect of their known range, searching for sighting, spoor (tracks) and droppings, and estimated that between 1,000 and 1,100 animals were alive. One of the largest populations was then at Wolong (130–150 animals). George Schaller (New York Zoological Society) considered the estimates to be fairly accurate based on his own observations (Schaller,



Baby and mother Li Li at He Tau Ping Research and Conservation Centre for the giant panda, Wolong Natural Reserve in October 1986 (Janice Martenson).

G.B., Hu Jinchu, Pan Wenshi & Zu Jing *The Giant Pandas of Wolong Univ.* Chicago Press, 1985).

By 1983 the Chinese government had established 12 panda reserves holding about 60 per cent of the wild panda population. Hunting of pandas was outlawed and a 2-year jail sentence was established for the killing of a giant panda. At least one farmer who accidentally snared a giant panda in the Wolong Reserve was imprisoned, and the fear of such penalties has made Chinese scientists and conservationists reluctant to handle the animals, despite assurances of support by the Ministry of Forestry in the case of accidental loss. At the Wolong Reserve, a giant panda conservation, research and breeding station was completed in 1984 and 11 captive pandas are housed there.

A new census commissioned by the national Sichuan provincial and local forestry bureau showed that by autumn 1985 several populations, including that of the large Wolong reserve, had declined by as much as 30 per cent since the mid-1970s (Ken Johnson, WWF and University of Tennessee, Knoxville). Extrapolation to other reserves suggests that the final number of free-ranging pandas may be as few as 600–700 animals. The remaining six mountain range habitats seem to be further subdivided by impenetrable barriers (for example, rivers, high ridges, logging roads and clear-cut forest) which fragment the pandas into small island populations. MacKinnon and Johnson estimate that about 35 such populations exist and that most have fewer than 20 individuals. Additionally, 10–20 per cent of the isolated panda populations found in the 1970s have disappeared.

Why are numbers of free-ranging pandas declining despite government protection? There seem to be three main reasons. First, and most critical, is human encroachment. The ratio of people to pandas in China is more than one million to one, resulting in considerable pressure on the remaining sources of timber and land suitable for agriculture. Although the panda reserves contain 60 per cent of the remaining wild animals, thousands of people also live there. Giant pandas are extremely timid and will give up habitat within miles of a human settlement. Settlements in Sichuan's mountains are growing: in Wolong Reserve there are more than 3,000 people and about 100 pandas.

The second factor is forest development. The Ministry of Forestry manages the reserve both for pandas and for timber supply. Satellite photos of panda habitat in China during the past decade show a direct correlation between forest harvesting and destruction of suitable panda habitation (de Wulf). As pandas need a greater than 70 per cent forest cover over the dense bamboo, certain types of forestry operations have in many cases effect-

ively eliminated that forest canopy.

The third factor arises from the periodic (about every 40 years) flowering and die-off of bamboo species. Although widely held to be a serious threat to the panda, it now seems that the flowering episodes are natural ecological events that the giant panda, as a species, has come to accommodate. The problem is critical only in the context of the habitat destruction and in association with the first two factors. Pandas can survive on secondary bamboo in the habitat when the principal species flowers, or if none is available they can migrate to an area where the same species has not flowered. As a result of the flowering of arrow bamboo in Wolong in 1983–85, 40–50 per cent of its pandas may have died and others may have emigrated.

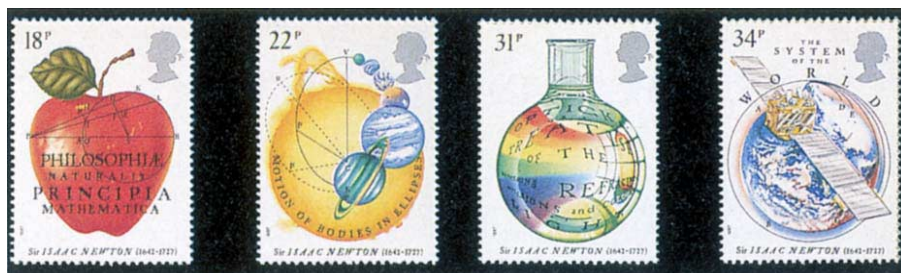
The WWF plan recognizes the pressures of human development on panda habitat and outlines a programme of relief, including a reassessment of reserve boundaries. Even if the habitat were restored, however, the precarious nature of the free-ranging panda is now a numbers game. As numbers drop and small populations become isolated, two factors



Captive panda Hsing Hsing at National Zoological Park, Washington DC (Jesse Cohen).

contribute to species demise. The first is demographic crashes: when small numbers of individuals constitute a population, the next generation may not reproduce simply because of chance effects such as accidents, altered sex ratios, snaring and bamboo flowering. A second concern is that small populations will inbreed by necessity. The genetic consequences of such 'founder effects' include inbreeding depression by expression of deleterious recessive loci that can affect reproduction, fecundity and survival. In addition, genetic homogenization by forced inbreeding removes endemic genetic variability with its evolutionary insurance against viral, bacterial and parasitic epizootics that afflict species in the wild. Overly inbred populations have a genetic 'axe' suspended over them in the form of extreme vulnerability to pathogens in the environment. These severe genetic consequences have not yet been demonstrated in wild pandas.

Conservation-genetic theory makes



THESE stamps mark the 300th anniversary of the publication of Isaac Newton's book *Philosophiæ Naturalis Principia Mathematica* and represent aspects of the author's work on gravitation and optics. The stamps are designed by Sarah Godwin and will be issued in the United Kingdom on 24 March. □

some predictions about small populations which are relevant to the panda's crisis. Michael Soule (*Zoo Biol.* 5, 101–114; 1986) estimates that to have a viable population, a principal aim is to retain 90 per cent of the endemic genetic variation for more than 200 years. Empirically, animal breeders find that inbreeding depression can be avoided if the rate of inbreeding per generation remains below 1 per cent, which translates into an effective breeding size (the number of bred animals in a population) of 50. This number is a minimum estimate because the logic assumes that the starting population was sufficiently outbred initially, an unlikely assumption with pandas, some of which have a reduced amount of allozyme variations (see O'Brien, S.J. *et al. Science* 223, 1127–1129; 1984). Further, because most of the 35 island populations have less than 20 individuals, few of the remaining panda isolates qualify as a 'viable population' in the context of predictive genetics. For this reason a possible strategy would be to re-establish the migration corridors between isolated populations to create larger panmictic wild populations. The extent of migration need not be great (less than 1 per cent per generation) to achieve enough gene-flow exchange to buffer the apparently inevitable consequences of the small population size of pandas with a very low rate of reproductive increase.

Given the low numbers of free-ranging pandas, the role of the captive breeding programme is very important to the preservation effort. Outside China, the captive population consists of about 14 animals, but inside the country there are about 100 captive giant pandas. Therefore, 10–20 per cent of the pandas live in captivity and with the rapidly diminishing wild populations, breeding these animals may effectively become pivotal in their survival. There are 137 Chinese zoos (Mr Li, Director of the Beijing Zoo), most of which want breeding pairs. The WWF management plan would improve captive breeding in conjunction with a reintroduction plan designed to reduce founder effects in the wild population.

The potential inbreeding problems perhaps caused by small founder popula-

tions become apparent from analysis of captive births of giant pandas. Of 51 captive births in China between 1963 and 1983 at the Beijing Zoo, only 19 cubs (37 per cent) survived for more than 2 months (Schaller *et al.*). This degree of infant mortality, a frequent consequence of inbreeding, is high compared with other zoo-bred species. Of the 51 births, 14 litters (64 per cent) were twins: pandas invariably raise only one member of a litter. Even if these 14 twins are considered as a single birth, the incidence of infant mortality is still 49 per cent.

Another concern is that all the recent captive litters of giant pandas in China are derived from 3 males and 10 females, an effective breeding size in captivity of 9.2 animals. By even the most optimistic of plans, the giant panda breeds worse than most species bred in captivity. But the small size of the effective breeding population may be improved by a coordinated management effort for the captive animals in China. Low fecundity may be a soluble management problem, because 8 of the 10 adults outside China have bred (Washington, Mexico City, Tokyo and a London–Madrid match). Artificial insemination has been successful in several cases and is often part of breeding programmes in China. Hence, Prince Philip's first query at Wolong was directed towards the initiation and progress of a cooperative management programme among Chinese zoological institutions.

The way forward for the giant panda must emphasize the importance of re-establishing panda habitat and implementing a coordinated captive breeding programme. There is cause for optimism because there is now a clearer perception of the problem and the solution seems attainable; most important, the Chinese people and government recognize the same hope and are determined to preserve the species. □

Stephen J. O'Brien is in the Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, Maryland 21701-1013, USA; and John A. Knight is Manager of the Doha Zoological Gardens, Qatar, on secondment from the Zoological Society of London, London NW1 4RY, UK.

Organic chemistry

Ring-current effects in C_{60}

R.B. Mallion

ONE of the most aesthetically appealing molecules to be diagnosed in recent years is buckminsterfullerene (Fig. 1), a carbon cluster — more dispassionately called icosahedral C_{60} — reported¹ two years ago by Smalley and co-workers. This apparently conjugated system (also, aptly, termed soccerballene and footballene) was independently suggested by two groups^{1,2} to be the first example of a spheroidal 'aromatic' molecule; π -electron and total-energy calculations have been consistent with this idea²⁻⁴. Smalley's group also proposed¹ that buckminsterfullerene should exhibit strong π -electron 'ring-current' effects⁵, and it is this question that is addressed by V. Elser and R.C. Haddon on page 792 of this issue⁶. These authors claim that the magnetic properties of icosahedral C_{60} are unlike those of any molecule yet encountered.

The ring-current idea of Ehrenfest, Pauling, Lonsdale and London, dating back to the 1920s and 30s, was revived with the advent of nuclear magnetic resonance (NMR) spectroscopy in the 1950s and, by the late 1960s, was the subject of considerable controversy⁵. These controversial

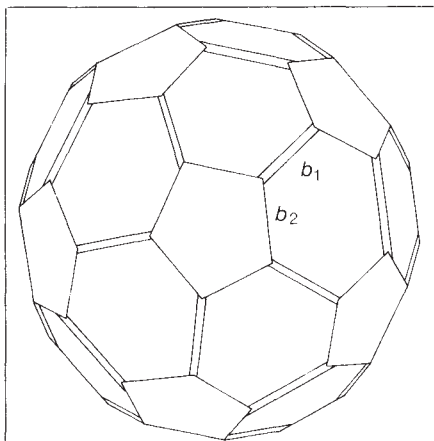


Fig. 1 Buckminsterfullerene (icosahedral C_{60} carbon cluster) and the two types of non-equivalent carbon-carbon bonds featuring in it: b_1 (shared between two six-membered rings); and b_2 (shared between a five- and a six-membered ring).

aspects are not often referred to these days, and the ring-current concept has to a large extent been rehabilitated⁷. Using a modification⁶ of London's original method that enables the Hückel molecular-orbital theory to be applied in a three-dimensional way, Elser and Haddon⁶ calculate that, contrary to the expectations of Smalley's group¹, the contributions to the magnetic susceptibility of icosahedral C_{60} from the 'mobile' π -electrons in this molecule are, overall,

only weakly diamagnetic.

All 60 carbon atoms in icosahedral C_{60} are equivalent by symmetry, but there are two distinct types of bonds: those shared by two six-membered rings (denoted by the symbol b_1 in Fig. 1) and those situated in both a five-membered and a six-membered ring (b_2). If β_1 and β_2 represent the respective resonance-integrals appropriate to these two types of bonds, then, from bond-length considerations, the ratio β_1/β_2 ($=k$) should be greater than one. On what they concede is the unlikely assumption that $\beta_1 = \beta_2$, Elser and Haddon calculate that the overall ring-current magnetic susceptibility of icosahedral C_{60} is only about one-fifth that of benzene and, furthermore, it is paramagnetic, rather than diamagnetic.

In these calculations, Elser and Haddon take account of the fact that, in experimental measurements of this kind, the direction of the external magnetic field is fixed in laboratory coordinates, whereas the molecule in solution is moving randomly. The NMR chemical shift (or magnetic susceptibility) is, of course, a tensor quantity, but because a molecule in solution is subject to this random motion, what is actually measured is a scalar quantity, averaged over all orientations of the molecule with respect to the applied, external magnetic field. The required average is the arithmetic mean of the three principal elements of the tensor. For benzene, this amounts to taking one-third of the 'static' value (calculated with the external magnetic field assumed to be normal to the plane of the molecule); buckminsterfullerene is, however, isotropic, and the rotationally averaged value is therefore identical with the static value.

Elser and Haddon point out⁶ that the assumption of equal resonance-integrals for the two types of bonds (that is, $\beta_1 = \beta_2$, implying that $k = 1$) is unrealistic; it is much more likely that $1.0 < k < 1.3$. But they show that, as the value assumed for k is varied, the π -electron ring-current magnetic susceptibility of buckminsterfullerene, predicted to be weakly paramagnetic for $k = 1$, changes sign near k values of 1.02. It is likely that k is considerably more than this and hence the authors conclude that because just a 2 per cent change in the assumed relative bond-strengths is sufficient to tip the balance between paramagnetism and diamagnetism, the π -electron ring-current contribution to the magnetic susceptibility of icosahedral C_{60} is very probably weakly diamagnetic; an atom placed at the centre

of the spheroidal system is, moreover, predicted by a non-quantum-mechanical method to be deshielded to the extent of about 0.6 parts per million in a hypothetical NMR experiment⁶.

Detailed examination of the contributions from individual molecular orbitals shows⁶ that this overall weak diamagnetism is in fact the result of fine cancellation of large paramagnetic contributions from some orbitals, and diamagnetic ones of similar magnitude from others. Such examination also reveals that C_{60}^{6-} is predicted to be strongly diamagnetic, and C_{60}^{12-} weakly diamagnetic — exactly the opposite of the pattern encountered with graphite and most planar, conjugated hydrocarbons, the ring-current magnetic susceptibilities of which tend to be diamagnetic for the neutral molecules and strongly paramagnetic for the corresponding reduced species.

The phenomenon encountered by Elser and Haddon⁶ — that the paramagnetic/

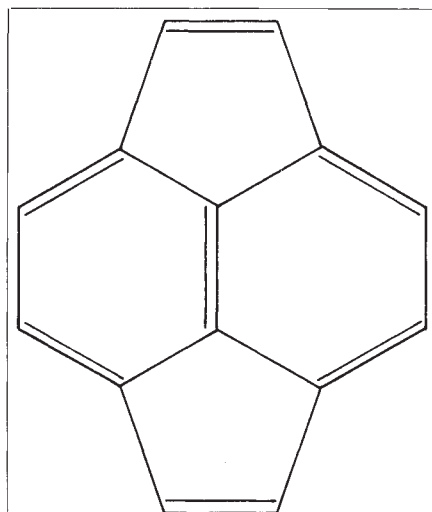


Fig. 2 Structural formula (carbon atoms only) of pyracylene, $C_{14}H_8$.

diamagnetic nature of the predicted magnetic susceptibility is dependent on the values assumed for the bond resonance-integrals — has its echo in observations made by myself and collaborators⁸⁻¹⁰ in earlier calculations on pyracylene (Fig. 2). Pyracylene, a planar hydrocarbon, has a carbon-atom skeleton formally to be found on appropriate parts of the surface of the buckminsterfullerene sphere. In molecules such as this, paramagnetic contributions will be the largest when magnetic-dipole transitions can take place between the ground-state (occupied) and excited-state (unoccupied) orbitals, and this activity will be particularly favoured when the energy separation between such orbitals is small.

The smaller this separation turns out to be, however, the more likely is any estimated value of it to be sensitive to the peculiarities of the specific method used for its calculation. Hence, the predicted magnetic properties of such molecules are

extremely dependent on the sophistication of the method (and, especially, of the wavefunction) used for their calculation.

If this (important) reservation is borne in mind, Elser and Haddon's theoretical predictions on buckminsterfullerene give a genuine impetus to future experimental work on the magnetic properties of this remarkable molecule. □

R.B. Mallion is Assistant Manager in Mathematics at The King's School, Canterbury CT1 2ES, UK.

Air pollution

Tracers of atmospheric transport

Hiram Levy II

ON page 800 of this issue¹, Gagosian and colleagues combine measurements of minute quantities of plant lipids present in organic aerosols collected over New Zealand with isentropic trajectory calculations to identify specific terrestrial source regions in the Southern Hemisphere and to determine transport paths over the southern oceans. Although the sensitivity of the chemical analysis and the ability to calculate believable 10-day trajectories are impressive, the key message of the paper is the value of using a combination of chemical and meteorological tracer techniques. Taken separately, neither approach would have produced conclusive results. In combination, the assignments of source regions and transport paths are quite powerful. Therefore, it is worth asking whether such techniques, which have been used successfully in relatively remote regions to identify distant sources, can provide quantitative information about global geochemical cycles and ultimately be applied to important local and regional environmental issues such as air pollution and acid rain in Europe and North America. Before discussing this question, an examination of some past developments may be useful.

Measurement techniques for radionuclides found on aerosols were first developed to assess movement of nuclear bomb debris and were found to provide valuable information on atmospheric transport. These techniques were then applied to ²¹⁰Pb and ⁷Be, the naturally occurring tracers of surface and stratospheric air². Next, filter measurements³ in the supposedly pristine air over the Caribbean uncovered episodes of very high levels of mineral aerosol (dust). Extensive measurements⁴ of dust over the remote

1. Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F. & Smalley, R.E. *Nature* **318**, 162–163 (1985).
2. Haymet, A.D.J. *Chem. Phys. Lett.* **122**, 421–424 (1985).
3. Klein, D.J., Schmalz, T.G., Hite, G.E. & Seitz, W.A. *J. Am. Chem. Soc.* **108**, 1301–1302 (1986).
4. Haddon, R.C., Brus, L.E. & Raghavachari, K. *Chem. Phys. Lett.* **125**, 459–464 (1986).
5. Haigh, C.W. & Mallion, R.B. *Progr. NMR Spectrosc.* **13**, 303–344 (1979).
6. Elser, V. & Haddon, R.C. *Nature* **325**, 792–795 (1987).
7. Gomes, J.A.N.F. *Molec. Phys.* **40**, 765–769 (1980).
8. Coulson, C.A. & Mallion, R.B. *J. Am. Chem. Soc.* **98**, 592–598 (1976).
9. Mallion, R.B. *Pure Appl. Chem.* **52**, 1541–1548 (1980).
10. Gomes, J.A.N.F. & Mallion, R.B. *J. org. Chem.* **46**, 719–727 (1981).

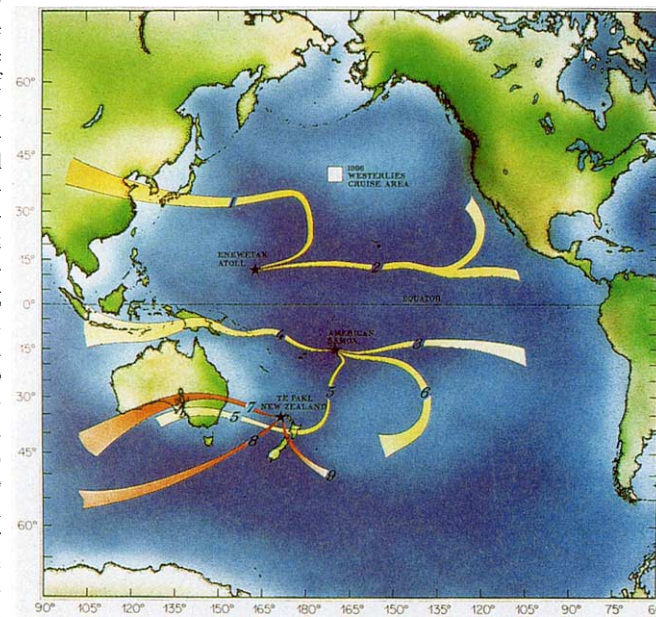
oceans then identified large active continental sources for the material fraction of oceanic sediments. Trace metals, which were found in minute quantities on aerosols in clean air and were detected by neutron-activation analysis, as well as other sensitive analytical methods, provided another set of chemical tracers. These were used to identify the source types for background aerosols and metals

provide only the horizontal component of the wind and the record itself becomes very sparse in remote regions. Near the surface, the observed history of the horizontal component can be used to reconstruct the path of the parcel, but such trajectories are not appropriate for transport over long distances, particularly when rising and sinking motions are involved. The first successful calculations of realistic long-range trajectories circumvented the lack of measured vertical velocities by assuming that the atmospheric motions were adiabatic (no heat supplied to or withdrawn from the air parcel). Available meteorological soundings were used to construct isentropic (adiabatic) surfaces on which the air parcels moved. The resulting trajectories identified transport paths of radionuclides and ozone carried down from the stratosphere to the surface of the Earth⁵. This approximation to full three-dimensional trajectories is quite accurate for large-scale motions in relatively dry and stable air, but it does not work where there is strong mixing, such as in the boundary layer or during convection, or when there is significant latent heat release from the condensation of water. Isentropic trajectories have since been used to identify tropospheric

sources; for example to track dust observed over the tropical Pacific back over 10 days to its source in arid central Asia⁶ and, as reported in this issue¹, to track plant waxes observed over New Zealand.

Although previous studies have identified source regions and transport paths, they have produced little, if any, quantitative information. To do so, chemical transformations and physical processes such as aerosol coagulation and settling, dry deposition and precipitation removal must be included, either directly in the trajectory calculations or indirectly via an appropriate surrogate. Furthermore, it is not sufficient to do this for a single event. The detailed study of one or more representative transport events is important for testing a particular technique, be it chemical or meteorological,

but a method is needed that integrates the results of all the events in a year or, given the year-to-year variability of weather, perhaps in a decade. As a first step, correlating surface wind direction, a primitive meteorological trajectory, with the deposition of acid rain in Bermuda clearly identifies North America as the primary source¹⁰. Without including chemical transformations and physical loss processes, however, the size of the North American export source cannot be determined.



Map of the overall air mass trajectories at the major field sites of the SEAREX project, including the station used by Gagosian *et al.*¹.

such as cadmium and lead⁵ and made very sensitive tracers of automobile exhaust⁶. The recent ability to measure small quantities of organic compounds found in aerosols⁷ provided biomarkers that not only differentiate among terrestrial, oceanic and anthropogenic sources, but identify specific regions.

In theory, the meteorological record of winds measured every 12 hours can be used to trace back in time the three-dimensional path of an air parcel, but in practice the normal meteorological data

The chemical tracer and isentropic trajectory methods of Gagosian and colleagues¹ seem to have some potential for quantitative studies of global geochemistry, particularly when there are a few well-separated sources and the background concentrations are low. But this type of approach is not adequate for studying regional phenomena, such as acid rain, which involve complex meteorology, chemical transformations, precipitation removal and several nearby large sources. Assigning the blame among all the neighbouring sources for acid deposition in one country or region is particularly difficult. A statistical approach using an array of chemical tracers with different signatures for different sources has been proposed by Rahn¹¹ and used, with some success, to apportion the sulphur fraction of acid rain in the northeastern United States among the different regional sources. However, this technique is not yet quantitative and does not work for the nitrogen fraction of acid rain. The

most likely approach to pinpoint the sources of atmospheric pollution, although guided by chemical tracers, will require the development of high-resolution chemical transport models driven by realistic simulations of the meteorology of a region. □

1. Gagosian, R.B., Peltzer, E.T. & Merrill, J.T. *Nature* **325**, 800-803 (1987).
2. Junge, C.E. *Air Chemistry and Radioactivity* (Academic, New York, 1963).
3. Delany, A.C. *et al. Geochim. cosmochim. Acta* **31**, 885-909 (1967).
4. Prospero, J.M. in *The Ocean Lithosphere, The Sea* (ed. Emiliani, C.) 801-874 (Wiley, New York, 1981).
5. Duce, R.A., Hoffman, G.L. & Zoller, W.H. *Science* **187**, 59-61 (1975).
6. Patterson, C.C. *Archs envir. Hlth* **11**, 344-360 (1965).
7. Simoneit, B.R.T., Chester, R. & Eglinton, G. *Nature* **267**, 682-685 (1977).
8. Danielsen, E.F. *J. atmos. Sci.* **25**, 502-518 (1968).
9. Merrill, J.T., Bleck, R. & Avila, L. *J. geophys. Res.* **90**, 12927-12936 (1985).
10. Jickells, T., Knap, A., Church, T.M., Galloway, J.N. & Miller, J.M. *Nature* **297**, 55-57 (1981).
11. Rahn, K.A. & Lowenthal, D.H. *Science* **228**, 275-284 (1985).

Hiram Levy II is at the Geophysical Fluid Dynamics Laboratory, Princeton University, Post Box 308, Princeton, New Jersey 08542, USA.

Developmental biology

Mechanism of fertilization

Brian Dale

RANKING high at present in the public eye is human *in vitro* fertilization. While politicians juggle with ethics on the use of human embryos in the laboratory, biologists have made some headway in understanding the mechanism of fertilization, using invertebrate gametes, that will help future clinical studies. Two problems of current interest are how the spermatozoon triggers the quiescent egg into metabolic activity; and why, of all the spermatozoa attached to the egg surface (Fig. 1), only one succeeds in penetrating the egg. Work recently published^{1,2} in the journal *Developmental Biology*, using techniques of electrophysiology and fluorescence microscopy to study the initial events of sperm-egg interaction in the sea urchin, have produced some unexpected results that are changing present concepts about fertilization.

The first of these papers, a collaborative effort between Frank Longo and Edward Chambers's group at Miami, correlates for the first time the sequence of ultrastructural events at fertilization with changes in the electrical properties of the egg plasma membrane. In a laborious study, the authors voltage-clamped eggs, fertilized them and then serially sectioned them to locate the fertilizing spermatozoon. The moment when sperm attach

to the egg was called time 0, and the first electrical event was detected 2 seconds later. Gamete fusion occurred at 7 seconds and cortical exocytosis, which leads to the elevation of a protective extracellular coat (see Fig. 2), was seen at 9 seconds. A second, larger change in conductance, indicating the global transmission of the activating impulse, was detected at 13 seconds. Chambers and McCulloch have taken the work further³, using the electrical measurement of capacitance to time sperm-egg fusion, thereby

reducing the time delay inherent in the fixation of biological material. It now seems⁴ that the initial electrical event occurs simultaneously with sperm-egg fusion and that in fact it may be the result of the ion channels in the sperm plasma membrane being incorporated into the egg plasma membrane.

In the second of the two recent papers², a group at Miami, led by Robert Hinkley, loaded sea urchin eggs with a fluorochrome and then fertilized them. When cytoplasmic continuity between the two gametes is established the dye flows into the spermatozoon, which then fluoresces. This ingenious way of timing sperm-egg fusion not only supports the results of Chambers and colleagues, but shows that when the egg membrane potential is voltage-clamped at a negative potential, only one sperm fuses with the egg. This is unexpected, because it was thought by many that the positive shift in membrane potential at fertilization is a fast block to polyspermy⁵. Preventing voltage change by clamping should have resulted in multiple sperm fusions and entries.

There are two generally accepted ideas about how the spermatozoon triggers the dormant egg, a cell one million times its own volume, into metabolic activity. One⁶ is that binding of the spermatozoon to externally located receptors on the egg surface is the primary signal, which is then transduced across the membrane to the cell interior, mobilizing second messengers such as inositol trisphosphate and calcium ions. The other⁷ is that the spermatozoon contains a factor, soluble in the egg cytoplasm, that is released into the egg following gamete fusion, and which directly triggers the mobilization of secondary messengers. If the initial electrical event is the result of the spermatozoon fusing with the egg, as now suggested by Chambers and colleagues, it seems less

likely that an externally located receptor mechanism is responsible for egg activation. The idea that spermatozoa contain a soluble activation factor is not new: it was popular more than 70 years ago, before the concepts of cell surface receptors were conceived. This is not to say, of course, that surface receptors are not important in fertilization, but rather their role is in the recognition and binding of the gametes, not in physiological activation of the egg.

One of the most astounding facts to an observer watching the process of fertilization for the first time is why only one spermatozoon, of potentially hundreds investing the egg surface, manages to enter the egg (Fig. 1). In many animals, including mammals, eggs *in situ* do not

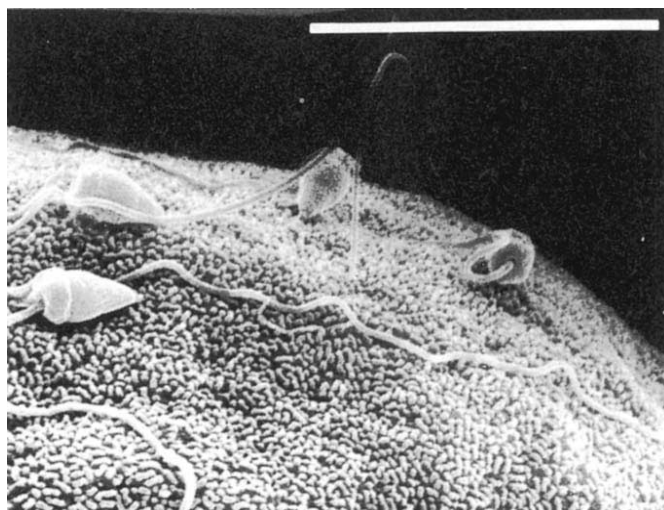


Fig. 1 The surface of a sea urchin egg a few seconds after insemination. Hundreds of spermatozoa can attach to the surface but only one succeeds in penetrating the egg. The successful spermatozoon dramatically alters the physiology of the egg within seconds. Bar, 10µm.

usually encounter large numbers of sperm, and in others the extracellular coats around the egg form a physical barrier. Nonetheless, in the laboratory, eggs challenged by hundreds of spermatozoa are usually monospermic. Several seconds after insemination, eggs from most animals undergo a complete structural reorganization of their surface, called the cortical reaction, that makes the egg refractory to further insemination (Fig. 2). But the cortical reaction is a slow propagating wave that requires many seconds or minutes, depending on the species, to spread over the surface. Thus the mechanism of fertilization must allow one sperm to interact with the egg and block the progression of supernumerary sperm within a second. An electrical fast block was considered appropriate until recently, when voltage-clamp and current-clamp experiments^{8,9} cast doubt on the hypothesis. If a positive voltage prevents polyspermy, then eggs kept at negative voltages during fertilization should be polyspermic. As confirmed elegantly by Hinkley and colleagues², this is not so and, therefore, although voltage may be involved in

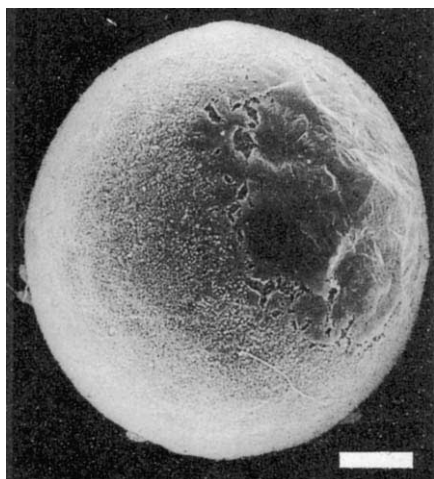


Fig. 2 The cortical reaction, the first morphological sign of activation, which involves the elevation of an extracellular coat around the egg. There, the coat has started to elevate at the top right-hand-side of the egg and will take a further 20 seconds to spread completely around the surface. Bar, 10 μ m.

sperm-egg fusion, it does not explain prevention of polyspermy. It will not be easy to think of another mechanism that changes the surface properties of a 100- μ m diameter cell in fractions of a second. The only other explanation at present is that the egg surface is not homogeneous as previously thought, but contains hotspots, pre-determined sperm-entry sites.

Research on human material is necessary to sustain and eventually to improve the success rate of a now-established treatment for some forms of sterility, and techniques to approach some of the major problems are available. Micro-injection of spermatozoa into close proximity to the egg plasma membrane, for example, may

be instrumental in cases of oligospermy (low sperm count) and unexplained infertility, and electrophysiological recording will help to classify stages of egg maturity and embryo quality. Progress in this field will undoubtedly benefit from what is known and what can be learnt from the gametes of invertebrates. $\square$

Brian Dale is at the Stazione Zoologica, Villa Comunale, 80121 Napoli, Italy.

Protein stabilization

Thermodynamics of unfolding

Walter Kauzmann

IT is widely believed that a major factor in maintaining the specific structures of native proteins is the tendency of non-polar side chains (such as those of the leucines, valine and phenylalanine) to pack together in the interior of the protein molecule where they can avoid contact with water. Thus, proteins are stabilized by the same physical forces as those that keep oil and water from mixing — a concept called the 'hydrophobic interaction'. There is, however, some controversy over the best way in which to approach this phenomenon. A recent paper by Robert Baldwin (*Proc. natn. Acad. Sci. U.S.A.* **83**, 8069–8072; 1986) shows, somewhat surprisingly, that the simplest model system of all, provided by data on the solubility of simple liquid hydrocarbons in water, can account quantitatively for some interesting regularities contained in the mass of thermodynamic data on protein-unfolding equilibria provided by the Soviet calorimetrist P.L. Privalov and others (see Privalov, P.L. *Adv. Protein Chem.* **33**, 167–240; 1979 for a review).

One of the most striking features of the thermodynamics of the unfolding of native proteins (a process that used to be called protein denaturation) is the existence of an enormous difference in the heat capacities of the folded (native) and unfolded (denatured) states. This difference amounts to 0.1–0.2 cal deg⁻¹ g⁻¹ protein, and Privalov showed that it correlates with the content of non-polar residues in the interior of the protein in its folded state. Simple liquid hydrocarbons such as benzene, toluene and hexane also show a large increase in their apparent specific heats on solution in water. This amounts to 0.67 cal deg⁻¹ g⁻¹ for benzene and 1.2–1.3 cal deg⁻¹ g⁻¹ for pentane and hexane — roughly ten times larger than the change on protein unfolding, on a per gram basis. These numbers for hydrocarbon solubilization in water are astonishingly large and are similar to the specific heat of water, which, at 1.0 cal deg⁻¹ g⁻¹, is itself one of the largest specific heats known, being approached or exceeded

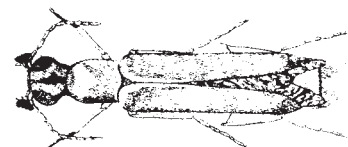
1. Longo, F. J., Lynn, J. W., McCulloch, D. H. & Chambers, E. L. *Devl Biol.* **118**, 155–166 (1986).
2. Hinkley, R. E., Wright, B. D. & Lynn, J. W. *Devl Biol.* **118**, 148–154 (1986).
3. McCulloch, D. H. & Chambers, E. L. *J. gen. Physiol.* (in the press).
4. Dale, B. & Santella, L. *J. Cell Sci.* **74**, 153–167 (1985).
5. Jaffe, L. A. *Nature* **261**, 68–71 (1976).
6. Turner, P. R., Jaffe, L. A. & Fein, A. *J. Cell Biol.* **102**, 70–76 (1986).
7. Dale, B., De Felice, L. J. & Ehrenstein, G. *Experientia* **41**, 1068–1070 (1985).
8. Shen, S. S. & Steinhardt, R. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1436–1439 (1984).
9. Dale, B. *J. exp. Biol.* **118**, 85–97 (1985).

only by such materials as solid lithium, liquid or gaseous helium, solid, liquid or gaseous hydrogen and, perhaps the most surprising of all, liquid ammonia.

Because of the large change in the heat capacity between the folded and unfolded states, the energy and entropy changes for protein unfolding and for the solution of hydrocarbons in water are strongly temperature-dependent. For the solution of simple liquid hydrocarbons in water, the enthalpy change, ΔH , becomes equal to zero at 15–30 °C. Thus, the solubility of benzene in water passes through a minimum at 16 °C because $\Delta H = 0$ at this temperature for the transfer of a molecule of benzene from liquid benzene to water. At this temperature the slight solubility of benzene in water entirely results from a large negative value of ΔS for the transfer, so the rejection of benzene by water at 16 °C is said to be entirely entropy driven. On the other hand, Baldwin points out that for all these simple liquid hydrocarbons ΔS for the transfer vanishes at a common temperature close to 113 °C, which means that the low solubility of benzene in water near the boiling point of water is mostly enthalpy driven.

Using thermochemical data obtained by J.M. Sturtevant (*Proc. natn. Acad. Sci. U.S.A.* **74**, 2236–2240; 1977) on lysozyme, Baldwin then shows that if the ΔH and ΔS of unfolding are partitioned between hydrophobic contributions (for which ΔH

THE ship-timber beetle *Lymexylon navale*, formally a serious threat to timber and wooden ships, forms part of the



recent revision of the Lymexylidae genera (Quentin D. Wheeler, *Bulletin of the American Museum of Natural History* **183**, 113–210; 1986). $\square$

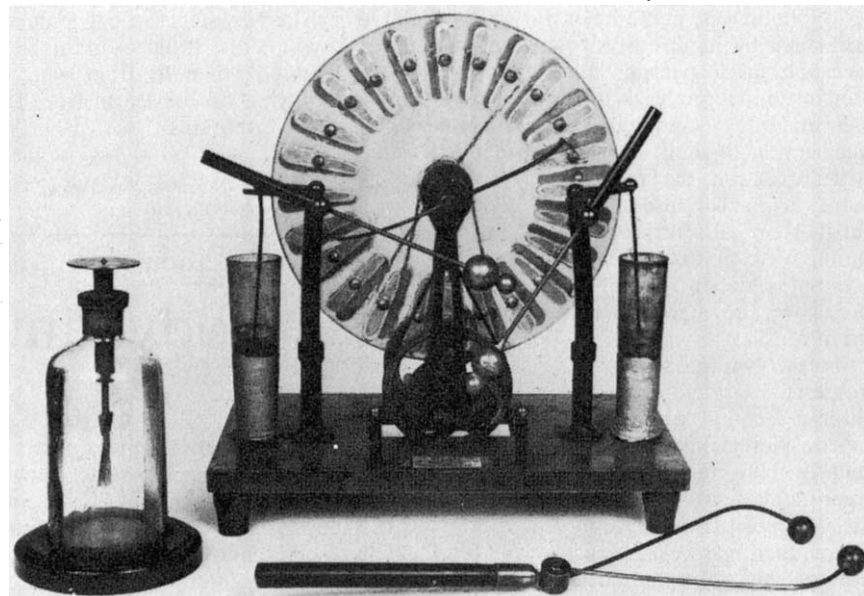
$=0$ at 22°C and $\Delta S = 0$ at 113°C) and residual contributions, ΔH_{res} and ΔS_{res} , then the latter turn out to be independent of the temperature. They presumably represent some contribution from the protein molecule itself apart from the hydrophobic interaction. Furthermore, both ΔH_{res} and ΔS_{res} are large and positive, as would be expected if the unfolded protein is anything like a random coil in which the favourable packing interactions that exist in the native state have been disrupted. Even more significantly, the per residue value of ΔS_{res} turns out to be about the same for several simple globular proteins. Privalov had already shown that the plots of temperature against ΔS for unfolding per residue for several typical small proteins tend to intersect near 110°C , which is a consequence of Baldwin's considerations.

Baldwin does not comment on the molecular basis of his observations; what he has to say is independent of that. This matter has been discussed at length by many authors and is controversial. I still believe that the Frank and Evans 'iceberg' model of 40 years ago is essentially correct: when it is transferred into water the non-polar hydrocarbon molecule induces, in the layer of water immediately surrounding it, a cage of more or less fully hydrogen-bonded water molecules (the iceberg). The large decrease in entropy that is known to accompany the transfer is accounted for by the increase in order accompanying the formation of the iceberg. As the temperature is raised the iceberg melts bit by bit — whence the large contribution to the specific heat. Indeed, this notion has been dealt with quantitatively by S.J. Gill *et al.* (*J. phys. Chem.* **89**, 3758–3761; 1985). This interpretation is also supported by molecular simulations of methane and neon in water, which invariably show a layer of structured water surrounding the non-polar solute molecule. Some of the attempts to explain these observations in other ways seem to me to be ludicrous, to say the least.

Baldwin's success with his model is remarkable in several respects. He wishes to explain the interaction with water of a non-polar group, R, in a molecule R-X, where X can be a single atom, such as a halogen (as in methyl chloride) or a group of atoms (as in ethanol or acetic acid, where X is $-\text{OH}$ or $-\text{CO}-\text{OH}$). To do this he studies the behaviour of the molecule R-H in water, as if the nature of the group X makes no difference to the interaction between water and R. That this will work at all is perhaps not remarkable if X contains only a few atoms, but in Baldwin's

THE picture of vortices on the cover of the 22–28 January issue of *Nature* was produced by Jim Salem (Thinking Machines Corporation, Cambridge, Massachusetts) and Stephen Wolfram (University of Illinois). □

Sale by Wheatstone Laboratory, London



A NINETEENTH century electrostatic generator (Wimshurst pattern) by Becker & Co., with a gold-leaf electroscope and a discharging fork. The Wimshurst machine can generate thousands of volts with a few turns of the handle, producing sparks between the electrodes. These items are included in a sale of scientific and philosophical apparatus by the Wheatstone Laboratory, King's College, London, at Christie's, South Kensington, London, on 5 March at 10.30 a.m. □

case X is the polypeptide backbone, with enormous steric and other influences that could easily interfere with the success of the approach. Evidently, they do not.

The point has also been made that the interior of a folded protein molecule is more likely to resemble a solid than a liquid, so that it is wrong to use the liquid hydrocarbon as the analogue with the clusters of non-polar residues that are known to exist in protein interiors. Baldwin's success shows that this cannot be an important consideration in the phenomenon that concerns him here.

But there is a fly in the ointment. The liquid hydrocarbon model fails almost completely when one attempts to extend it to the effects of pressure on protein folding. Just as temperature effects on the unfolding equilibrium are determined by the enthalpy change, ΔH , of the process, so the pressure effects are determined by the volume change, ΔV , that accompanies the unfolding. Studies on several proteins clearly show that ΔV for unfolding is positive at low pressures but becomes negative at pressures above 1–2 kbar. (This shows that the compressibilities of the folded and unfolded forms must be very different, just as their heat capacities are different.) On the other hand, studies of the effect of pressure on the solubility in water of non-polar substances just as clearly show the opposite behaviour: ΔV for the transfer from non-polar to aqueous environment is negative at low pressures but becomes positive at pressures above

1–2 kbar. This failure of the hydrophobic liquid model is mentioned in passing by Baldwin, and workers in the field are probably aware of it in a general way, but the fact is that this discrepancy could well be devastating to the whole approach. Volume and enthalpy changes are equally fundamental properties of the unfolding process, and no model can be considered acceptable unless it accounts for the entire thermodynamic behaviour.

It is, of course, much easier to study the effects of temperature on systems than to study pressure effects. Thermostatic baths are relatively cheap and very easy to work with, whereas high-pressure apparatus (I refer to pressures of up to 5 kbar, which is really not very high) is expensive and awkward to use. It is easy, therefore, to sweep this discrepancy under the rug. A great deal more work has been done on temperature effects, and there are many more data to work with. But one is reminded of the story of the drunk who, one dark night, has lost his keys. He is seen looking for them under a street light. When asked where he lost them, he points across the street, where all is dark. "Why, then, are you looking for them here?" He replies, "Because there is more light here". Until more searching is done in the darkness of high-pressure studies, our understanding of the hydrophobic effect must be considered quite incomplete. □

Walter Kauzmann is Emeritus Professor of Chemistry at Princeton University, Princeton, New Jersey 08544, USA.

The hidden dangers of AIDS vaccination

SIR—Vaccines against AIDS (acquired immune deficiency syndrome) will probably be directly tested in human immunodeficiency virus (HIV) seronegative humans before a thorough evaluation of their potential hazards is completed in animals¹ because of the shortage of chimpanzees^{1,2}, and the first vaccination-like programme in HIV-seropositive subjects is already being conducted in Africa³. But both types of experimentation may prove harmful to the volunteers even though the injected material is not infective.

Indeed, vaccination⁴ and post-infection immunization⁵ against other lentiviruses closely related to HIV⁶⁻⁹ have induced deleterious effects. Goats injected with inactivated caprine arthritis-encephalitis virus develop more severe lesions than controls when later challenged with the live virus⁴. And the induction of a greater severity of lesions in visna virus disease by post-infection immunization⁵ of sheep suggests the involvement of specific antigenic stimulation.

Autoimmune mechanisms may be involved in the production of the specific CD4 lymphopaenia of AIDS^{10,11}. Therefore, we cannot exclude the possibility that AIDS vaccination in HIV-seronegative subjects may increase the severity of disease if HIV is encountered later, especially if the vaccine is not fully protective. In the same way, immune stimulation of seropositive volunteers with HIV-derived products may accelerate the disease; however, antigenic stimulation of their CD4 cells may increase the propagation of viruses to uninfected cells^{10,11}.

Such untoward effects are less likely to be produced by viral antigens that induce a strong neutralizing antibody response, but the ability to induce neutralizing antibodies in humans cannot be fully predicted from experiments in animals.

The expansion and prognosis of AIDS will increase the pressure to test AIDS vaccination in humans. The volunteers in such programmes should of course be fully informed about the risks they take and be able to understand this information. But the data obtained so far in animal immunization against lentiviruses closely related to HIV are a strong indication that the routine testing of the safety of vaccines in animals before use in humans should not be abandoned.

AXEL ELLRODT
PHILIPPE LE BRAS

Service de médecine interne et INSERM
U56,
Centre Hospitalier Universitaire de
Bicêtre,
94275-Le Kremlin-Bicêtre, France

1. Newmark, P. *Nature* **324**, 304-305 (1986).
2. Barnes, D.M. *Science* **233**, 1149-1153 (1986).

3. Newmark, P. *Nature* **325**, 290 (1987).
4. McGuire, T.C. *et al. Am. J. vet. Res.* **47**, 537-540 (1986).
5. Nathanson, N. *et al. J. comp. Pathol.* **91**, 185-191 (1981).
6. Sonigo, P. *et al. Cell* **42**, 369-382 (1985).
7. Chiu, I.M. *et al. Nature* **317**, 366-368 (1985).
8. Gonda, M.A. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 4007-4011 (1986).
9. Gonda, M.A. *et al. Science* **227**, 173-177 (1985).
10. Klatzmann, D. & Montagnier, L. *Nature* **319**, 10 (1986).
11. Klatzmann, D. & Gluckman, J.C. *Immun. Today* **7**, 291-296 (1986).

A different perspective on coral population genetics

SIR—I wish to raise two objections to Diamond's News and Views article "Clones within a coral reef"¹. First, Diamond misrepresents the reproductive and juvenile ecology of *Porites compressa* by equating the rarity of small (less than 60 cm²) colonies with infrequent successful larval settlement. Low larval abundance and post-settlement mortality could also explain why small colonies are rare. There is no apparent shortage of larvae. Following one of several spawnings during a summer, the water can become cloudy with *P. compressa* eggs, and densities of 10 larvae m⁻³ sea water occur in Kaneohe Bay², the site of the study³ reported by Diamond¹.

Before settlement, *P. compressa* larvae are less than 0.3 mm long². A juvenile coral, under field conditions, is not visible until 2 mm in diameter⁴, for *P. compressa*, perhaps three months after settlement, extrapolating from growth of other species⁵. Larval settlement is also often cryptic^{4,7}. Thus, only careful examination under the microscope of substrata shortly after spawning can determine if unsuccessful larval settlement or post-settlement mortality explain the scarcity of small colonies. The latter is strongly supported by studies from the Great Barrier Reef. Examinations of substrata give much higher estimates of the number of corals successfully settling than do macroscopic observations^{4,6}, and larval mortality is highest during the first three months after settlement⁷. Regardless of why juveniles are rare, they are important, representing new genotypes in the population.

Second, Diamond fails to discuss potential problems with using morphological traits and ultraviolet-absorbing pigments

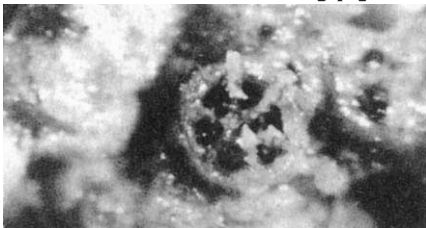


Fig. 1 *Porites compressa*, a recently settled juvenile, with the polyp retracted in the skeleton. Photographed on an experimental substratum collected from windward Coconut Island, Kaneohe Bay, Oahu, Hawaii. Maximum diameter is 0.45 mm. Juveniles of this size are not visible without magnification.

to identify genotypes. Both types of characters successfully separated genotypes in the small area of uniform depth³ investigated in the study Diamond reports. However, *P. compressa* and other species can exhibit both changed morphology⁸⁻¹⁰ and concentration of ultraviolet-absorbing pigments⁸ when transplanted to different habitats and depths. Before morphological characters and ultraviolet-absorbing pigments are used to distinguish genotypes in more heterogeneous environments, their phenotypic plasticity must be investigated further.

RACHEL FITZHARDINGE

Hawaii Institute of Marine Biology
and Zoology Department,
University of Hawaii,
2538 The Mall,
Honolulu, Hawaii 96822, USA

1. Diamond, J.M. *Nature* **323**, 109 (1986).
2. Hunter, C.L. *Mar. Ecol. Prog. Ser.* **26**, 67-71 (1985).
3. Hunter, C.L. *Proc. Fifth Int. Coral Reef Congr.* **6**, 69-74 (1985).
4. Wallace, C.C. *et al. Mar. Ecol. Prog. Ser.* **32**, 299-306 (1986).
5. Harriott, V.J. *Proc. Fifth Int. Coral Reef Congr.* **4**, 367-372 (1985).
6. Wallace, C.C. *Mar. Ecol. Prog. Ser.* **21**, 289-298 (1985).
7. Babcock, R.C. *Proc. Fifth Int. Coral Reef Congr.* **4**, 355-360 (1985).
8. Maraggs, J.E. thesis, Univ. Hawaii (1972).
9. Oliver, J.K. *et al. J. exp. Mar. Biol. Ecol.* **73**, 11-35 (1983).
10. Willis, B.L. *Proc. Fifth Int. Coral Reef Congr.* **4**, 107-112 (1985).

Is there such a thing as fractal music?

SIR—Philip Campbell¹ raises the possibility of music being derived purely from mathematical formulae — an acoustic analogy to the pictorial representations of arithmetical relations, such as the iterated mappings leading to 'fractal' images (Mandelbrot sets and the like). In brief, does fractal music exist and is it interesting?

The answer to both questions is yes². In fact, the fractal pieces that have been synthesized to date exhibit astounding paradoxes of perceived pitch³⁻⁵. One such 'composition' is a tone complex comprising many pure tones, spanning the auditory frequency range, whose adjacent frequencies are in the constant ratio 1:2^{1/12} (10Hz, 21Hz, ..., 428Hz, 907Hz, ..., 18,284Hz). The figure shows a plot of a waveform comprising seven such frequency components. Functions such as those (imperfectly) illustrated were introduced by Karl Weierstrass into the debates on the foundations of mathematics during the last century. (Specifically, Weierstrass wanted to show his sceptical colleagues that there are functions that are everywhere continuous but nowhere differentiable. In other words, what mathematicians call a function is a lot less innocent than they were wont to believe.)

If a (computer generated) tone complex with the above frequencies is recorded on magnetic tape and played back at twice

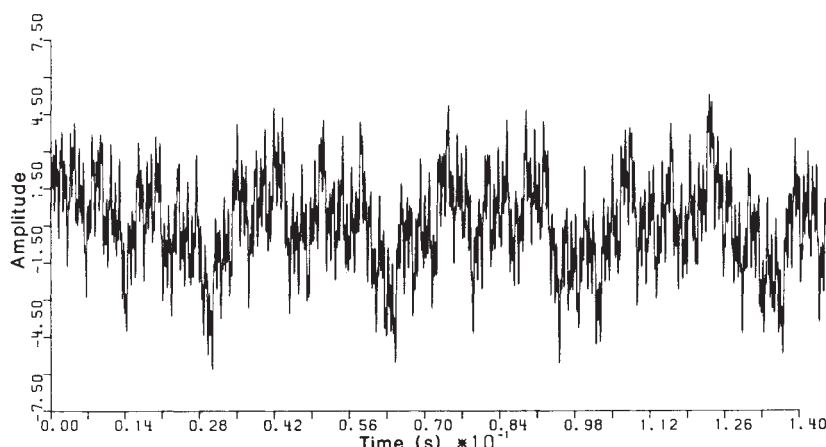


Fig. 1 Fractal waveform leading to paradoxical pitch perception.

the tape speed, it will sound a semitone lower in pitch rather than an octave higher (the usual percept engendered by tape speed doubling).

The explanation for this curious percept is based on the importance of adjacency in pitch perception^{6,7}. Thus, the tone component with a frequency of 428 Hz turns into one at 856 Hz upon doubling. But in comparing the two chords, the human auditory system will identify the frequency-doubled component at 856 Hz with the nearest component in the original complex, namely 907 Hz. And 856 Hz is a semitone lower than 907 Hz. The same is true for all other components and thus a lowered pitch will be perceived.

Self-similar fractal structures are applicable not only in the frequency ('pitch') domain but can also be introduced into the time ('rhythm') domain with interesting consequences. In fact, certain self-similar number-theoretic sequences, such as the binary Morse-Thue and 'rabbit' sequences², give rise to very appealing rhythms.

MANFRED R. SCHROEDER

Drittes Physikalisches Institut,
Universität Göttingen,
Bürgerstrasse 42-44,
D-3400 Göttingen, FRG

1. Campbell, P. *Nature* **324**, 523-528 (1986).
2. Schroeder, M.R. *Number Theory in Science and Communication. With Applications in Cryptography, Physics, Digital Information, Computing and Self-Similarity* 2nd edn Ch. 30 (Springer, New York 1986).
3. Risset, J.C. *Proc. 7th Int. Conf. Acoust., Budapest*, Vol. 3, 613-616 (Akadémiai-Kiadó, Budapest, 1971).
4. Charbonneau, G. & Risset, J.-C. *C. r. hebdo. Séance. Acad. Sci., Paris B281*, 289-292 (1975).
5. Schroeder, M.R. *J. acoust. Soc. Am.* **79**, 186-189 (1986).
6. Shepard, R.N. *J. acoust. Soc. Am.* **36**, 2346-2353 (1964).
7. Terhardt, E. *Hear. Res.* **1**, 155-182 (1978).

PHILIP CAMPBELL REPLIES — Musical quantities such as pitch and loudness have been found to exhibit $1/(\text{frequency})$ type power spectra (ignoring spectral peaks due to the basic rhythms of the pieces concerned), while noise synthesized with $1/f$ characteristics sounds more 'pleasant' or 'interesting' than Brownian ($1/f^2$) or completely random noise¹. In view of Dr Schroeder's comments it is interesting that

$1/f$ noise has fractal properties. Since writing my article I have learnt that at least one composer is investigating musical design using fractals².

1. Voss, R.F. & Clarke, J. *Nature* **258**, 317-318 (1975); *J. acoust. Soc. Am.* **63**, 258-263 (1978).
2. McNabb, M. in *The Language of Electroacoustic Music* (ed. Emmerson, S.) 150-153 (Macmillan, London, 1986).

The mechanism of chromogranin A processing

SIR—Two groups have recently conjectured that chromogranin A or a related protein may serve as a precursor for biologically active peptides in a number of cells of the diffuse neuroendocrine system^{1,2}, based upon the realization that pancreastatin, a peptide from porcine pancreas that inhibits insulin secretion, is closely related in sequence to residues 251-294 of bovine adrenal medullary chromogranin A³. What other evidence is there for this relationship?

Chromogranin A immunoreactivity in the pancreas is localized to a variety of endocrine cells of the islets of Langerhans⁴⁻⁶. The fact that pancreastatin as isolated has undergone C-terminal amidation suggests that it is produced in an intracellular, indeed intragranular, environment rather than being an artefact of tissue disruption. Enzymes which might excise pancreastatin from the chromogranin. A sequence through cleavage at single basic residues and which could catalyse its C-terminal amidation are certainly present in islets as such reactions occur in the processing of propancreatic polypeptide and proglucagon^{7,8}. Chromogranin A might also be processed at any of the eight sites of dibasic amino-acid sequence as endopeptidases with this specificity also exist in islet tissue.

In the adrenal medulla, where chromogranin A was first described as a protein of relative molecular mass (M_r) 72,000, there is little indication of major proteolysis of chromogranin A apart from N-terminal signal peptide removal. The half life of the protein in these cells is

comparable to that of the enkephalin precursor, being of the order of 6-8 days⁹. But a very different situation prevails in the pancreatic B-cell, where we first encountered chromogranin A as protein of M_r 21,000 (ref. 10). This protein, which we called betagranin, was cosecreted with insulin and, was shown to have an amino-acid composition and N-terminal sequence similar to bovine chromogranin A¹¹.

Pulse-chase labelling experiments in insulinoma cells show that betagranin is synthesized as a precursor indistinguishable in size and immunoreactivity from rat chromogranin A (M_r 80-100,000). This precursor undergoes rapid proteolysis soon after reaching the *trans* Golgi cisternae and entering the secretory granule compartment¹². The M_r 21,000 peptide appears to be a stable end product generated by initial cleavage of the chromogranin A-like precursor at a site equivalent to Lys114 Arg115 in the bovine sequence by the action of a Ca^{2+} -dependent acidic endopeptidase¹³. The C-terminal amino acids so exposed are removed by carboxypeptidase H. As the antiserum used in these experiments is directed towards the M_r 21,000 N-terminal fragment we can say little about the fate of the rest of the precursor apart from the fact that it too is processed but at a lower rate.

On a molar basis, the estimated amount of chromogranin-related peptides in insulin granules would at most represent one per cent of the insulin content¹⁴. As the circulating insulin level generally lies in the nanomolar range it would take a molecule with an exceptionally long half life to achieve a blood concentration approaching the 10 nM or so at which pancreastatin was shown to exert its effect *in vitro*. This might indicate that the chromogranin A-related peptides act locally on endocrine cells or adjacent vasculature. The concept that the pancreatic B cell responds to a product of its own secretion or to other islet hormones as part of an autoregulatory cycle has frequently been invoked. It would now appear that chromogranin A-derived peptides should also be seriously considered in this context.

J.C. HUTTON

H.W. DAVIDSON

M. PESHAVARIA

Department of Clinical Biochemistry,
University of Cambridge,
Cambridge CB2 2QR, UK

1. Eiden, L.E. *Nature* **325**, 301 (1987).
2. Huttner, W.B. & Benedum, U.M. *Nature* **325**, 305 (1987).
3. Tatemoto, K. *et al. Nature* **324**, 476 (1986).
4. O'Connor, D. *et al. Life Sci.* **33**, 1657 (1983).
5. Wilson, B.S. & Lloyd, R.V. *Am. J. Path.* **115**, 458 (1984).
6. Cohn, D.V. *et al. Endocrinol.* **114**, 1963 (1984).
7. Schwartz, T.W. *FEBS Lett.* **200**, 1 (1986).
8. Andrews, P.C. *et al. J. biol. Chem.* **261**, 8128 (1986).
9. Falkensammer, G. *et al. Neuroscience* **14**, 735 (1985).
10. Sopwith, A. *et al. Biochim. biophys. Acta.* **803**, 342 (1984).
11. Hutton, J.C. *et al. FEBS Lett.* **188**, 336 (1985).
12. Hutton, J.C. *et al. Biochem. J.* (in the press).
13. Hutton, J.C. *et al. Biochem. J.* (in the press).
14. Hutton, J.C. *et al. Endocrinology* (in the press).

A touch of (social) class

Chris Brand

Education and Class: The Irrelevance of IQ Genetic Studies. By Michel Schiff and Richard Lewontin. Clarendon: 1986. Pp. 243. £30, \$45.

Now that psychology has re-discovered cognition, and has lost interest in behaviourism and its laws of learning, general intelligence (*g*) has resumed its position as psychology's main theoretical construct — on a par with the gene for geneticists, with social class for sociologists and with the atom for chemists. This is not to say that psychologists can be relied on to accept *g* as originally enunciated by Charles Spearman: indeed, the aim of 'smashing' *g* is over-riding for many cognitive psychologists today. Yet *g* is certainly the concept of the widest explanatory potential known to psychology; it may possibly have a unitary biological basis — Volkmar Weiss's work in Leipzig implicates just a single gene locus as critical; and even the endeavours of cognitivists to break up *g* into psychological sub-components need to culminate in a story of how such sub-components can be glued together to yield the empirical observation that all measurable mental abilities are positively correlated with each other and predict socio-economic success.

To date, it cannot be said that cognitive psychologists have approached *g* with conspicuous clear-headedness. They tend to reach for the commonsense (and faintly behaviouristic) notion that, because strategies can be found for improving performance on most mental tests, high scores achieved without special opportunities for practice and learning can themselves be attributed to brighter testees behaving . . . well, more strategically: but cognitivists fail to specify what strategy, or even what particular combination of strategies, could be relevant to the enormous range of tasks at which the higher-*g* subject typically enjoys an advantage.

By contrast with such vagueness, the authors of *Education and Class* are a model of theoretical lucidity. For they at least can name the variable that does the work which the London school of psychology brazenly ascribes to *g* itself: the name of the variable is social class, to the praise of which they devote some masterly passages of propaganda (especially their Chapter 1) and many more problematical asides.

Nor is theory alone the strongpoint of

these authors. For the centrepiece of their book is a solid piece of natural experimentation (first reported in *Science* 200, 1503–1504; 1978) which must put British social scientists and their funding agencies to shame for neglecting such opportunities for empirical enquiry. With commendable assiduity and sensitivity, the former nuclear physicist, Michel Schiff, and his

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Children at school — what determines their intellectual potential? Inset: Charles Spearman, originator of the concept of general intelligence.

French colleagues managed to trace 32 adopted children whose own biological mothers had borne other children (typically half-siblings of the adoptees) whom they had reared themselves. The essence of the study is the comparison of the intellectual and educational progress of these two, partly matched groups of children who were brought up in very different environments. The adoptees went to homes that were (on average) $1\frac{1}{2}$ standard deviations high in social class (at around the 93rd percentile); and the 39 children who stayed with their own biological mothers had homes that were 1 standard deviation low in social class (around the 16th percentile). As the authors quite properly boast, no such extreme comparison of biologically matched children has previously been made.

What was the result? According to the version presented here, the adoptees had a mean IQ of 108.7, while the equivalent

figure for the 'biological', non-adopted children was 94.6; and in progress at school the differences were wider. Thus are the authors emboldened to assert the causal status of social class — and to disparage other workers for their neglect of Marxist understandings of the travails of society under the artificial inequities of capitalism — and to hope that "passion can ascend through meticulous science to theoretical advance".

Sadly for Schiff and Lewontin's pleasingly simple idea of the causal importance for children of their parents' socio-economic status (SES), many qualifications are in order — some of them admitted, though with much huffing and puffing, by the authors.

(i) The apparent 14-point difference in IQ reduces to 8 points on non-verbal IQ tests, that is on those tests that would normally be used to gauge a child's true intellectual potential when its social environment had been poor. (This result appears in the French team's other publications, but not in the book.)

(ii) The parents of the adopted children would have supplied not only their high-SES homes but also the extra nurturance ("wantedness", as the authors put it) that has often been noted by other researchers to increase child IQ. Yet, despite their 'rich' environments, the adopted children were still 10 points lower in IQ than classmates at their schools who were of similarly high parental SES.

(iii) The adopted children may have been selected (for example as more alert or responsive) in infancy, and thus of greater intellectual potential.

(iv) Arthur Jensen himself has long supposed that each standard deviation of parental SES is worth (as a causal influence) some 3.35 IQ points to a child. So the adoptees in the present study, on a mainstream hereditarian account, should have had a boost of $2.5 \times 3.35 = 8.4$ IQ points (plus an extra advantage from adoption alone). Thus it is quite unclear what the authors feel they are battling against.

Further problems arise with the rest of the authors' contentions that SES is really the key variable in intellectual development and achievement.

(v) The authors discover (as did Britain's Robbins Commission in 1963) that higher parental SES boosts children's chances of entering a university, even when child IQs are matched. Yet the reverse procedure, matching children for SES and examining the influence of IQ, shows IQ to have been markedly more

important than SES, as it was in Robbins's data.

(vi) The methodological problems of demonstrating the potency of SES in non-experimental data are quite unresolved by the authors: they do not even attempt path analysis. One particular source of their problems may even be that it is SES, rather than IQ, which is not readily and sensibly measurable: their own scheme assigns lower SES to antique dealers, TV announcers and farmers than to garage owners, taxi drivers and medical secretaries. Perhaps it is time for the champions of SES to do some sociometric homework?

(vii) Occasionally the authors and their contributors themselves allow that "the variance of IQ scores of children of a given social class is about 80 per cent of the total variance in the population", that "the combined genetic and non-genetic effects of social class only add up to 20 per cent of total variance in IQ scores", and even that "our methodology does not allow one to exclude the possibility of a genetic difference between social classes". Once more, such admissions make one wonder how they have managed to sustain their general enthusiasm for the endearingly simple theory of 'classism'.

The solution to this conundrum is to be found in the impassioned rejections of elitism, eugenicism and IQ in which the book abounds. Essentially, the authors' stance is that no proposition about IQ can be accepted unless it can immediately be rewritten in ideologically convenient form. Yet casual re-writing can make for a wider incoherence. So a statement such as 'IQ tests are invalid and meaningless and don't reflect intelligence' is all right; but so is 'IQ tests discriminate between the social classes'. Both are ideologically attractive propositions, and possibly true: but what then, exactly, do IQ tests measure so as to yield their countless correlations? The problems of such impetuosity for the reader mount as the pages go by. If the social classes differ in IQ in some study, then the tests are invalid (in some way); if the social classes are the same in another study, then the elitist testers are exposed as forlorn incompetents. Partial correlations can be used to demonstrate the causal efficacy of SES, but certainly not to do the same for IQ. In any case, all genetic studies, even the authors' own, are irrelevant to helping the working class. Most crucially, IQ testers are held to suffer from "brutal pessimism" about the human condition: yet the authors of the present plea for State-orchestrated equalization would strip society of just those artificial products of capitalism, the advantaged families that alone seem capable of providing 14- (or 8-) IQ-point boosts to the small proportion of working class children that could ever — even under the most Maoist arrangements — be adopted by them.

Amidst the confusion, some perfectly

reasonable points are made (especially in Chapter 5) in objection to the more preposterous straw men that can be made to lumber forward from hereditarian ranks. Yet the authors' insistence that the effect of genes depends on the environment would be more impressive if a single demonstrable example of this undoubted possibility could be produced from human psychogenetics: it is not clear why anyone should believe in 'reaction ranges' and 'interaction effects' as central to the causation of present human intellectual differences merely because such effects can be demonstrated by experimenters who can make gross alterations to the environment for batches of spring wheat. Moreover, the authors' assertion that "genes do not make eyes or brains" could be very misleading to those many social scientists and cognitive psychologists who already play down the degree to which humanity has a biological infrastructure as well as, for sure, a psychological and cultural superstructure.

Uncannily, as if sensing the need for more mystification with which to cover their erratic tracks, Schiff and Lewontin finally choose to obfuscate their historical materialism by insisting that "the human animal constructs itself". This propitious if somewhat fuzzy idea may well appeal to divines, to social psychologists and to the more desperate cognitivists; and it may serve as a reminder of Sandra Scarr's recent claims (unmentioned by these authors) that genetic influences express themselves more powerfully as growing children have more scope to select and create their own environments. Yet whether we will ever understand such 'self-construction' without also understanding even the most salient and well-attested features of our biology must be the first question for readers of this essentially flat-Earthist tract. □

Chris Brand is a Lecturer in the Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.

Stuff and sense

Robert W. Cahn

Advancing Materials Research. Edited by Peter A. Psaras and H. Dale Langford. *National Academy Press, Washington, DC: 1987. Pp.391. \$47.50. In Britain distributed by Wiley, about £45.*

In October 1985, a conference was held in Washington, DC, to mark the twenty-fifth anniversary of the creation of the government-financed Materials Research Laboratories (MRLs). Initially there were three of them — at Cornell, Northwestern and Pennsylvania Universities — but many more were set up later and over the years they have collectively exerted an enormous influence on materials science research in the United States. This book is a record of the meeting.

The opening papers are by William Baker, Robert Sproull and Lyle Schwartz, two of whom were midwives to the birth. They map out the individual pressures and institutional responses which led to the creation of the MRLs; describe their mid-stream transfer from the Advanced Research Projects Agency (ARPA) of the Defense Department to the National Science Foundation (NSF), and the strategic changes which this entailed; and finally attempt a judgement as to the success, past and present, of the whole enterprise. The interdisciplinary 'thrust groups' (or Materials Research Groups) favoured by the NSF to replace, progressively, the more substantial MRLs of earlier years, have focused on certain topics such as 'organic metals', ultralow temperatures, intercalation compounds and a whole

range of novel materials, which are surveyed by Lyle Schwartz: he claims that many of these activities were made possible by the cooperative organization pioneered by ARPA and refined by the NSF.

Schwartz's bird's-eye view of the thrust groups' activities leads on to eleven magisterial surveys of particular fields; here, very effective use is made of limited space to convey the flavour of recent work in metallurgy, condensed-matter physics in relation to materials, ceramics, organic polymers, materials synthesis and processing, chemistry as applied to materials science, and several others. Several of these overviews are extremely impressive, marrying broad coverage with an emphasis on the most recent developments that have especial potential for scientific growth. Ductile ordered alloys, heavy-electron compounds, reptation of long-chain polymers and ceramic packaging of microcircuits are just a few of the topics discussed. John Cahn and Denis Gratias contribute yet another survey of quasicrystals (a mature subject, yet barely three years old). In this, the main section of the book, one has the sense that for once achievement has been as impressive as promise — the entire 270 pages can be recommended to labourers in the materials science vineyard who seek surveys of the interdisciplinary varieties grafted on to the solid, disease-resistant rootstock of the central disciplines.

A final short section, by a number of authors from many industries and universities, discusses some of the hardy perennials of science policy. Some but by no means all of this section is restricted to purely American issues. Small science versus large science and the grave prob-

lems of getting equipment and financing support staff in the MRLs and MRGs are debated here, as is the expected role of particular families of materials and processes in particular industries (near-net-shape manufacture, ceramics in heat-engines, surface modification, molecular-beam epitaxy and so on). Several contributors make points which would risk provoking cynical shrugs in Britain today: Albert Clogston, late of Bell Laboratories, insists on the crucial importance of *basic* technology in research and development laboratories, while Praveen Chaudhari, a senior research manager at IBM, declares that "an important point which cannot be taken for granted or emphasized enough is that the research enterprise of the nation requires an infrastructure that nurtures general science, or science that cannot be identified at present with any particular area of application". This comes from an industrialist, not from an academic!

Earlier in the book, Sproull emphasizes how the MRLs were based on the post-war recognition that Government contracts and grants "should provide much more scope for the contractor's imagination and discovery and more harvesting by informal agency-contractor interactions rather than by fulfilling specifications", and that the Office of Naval Research (one of the progenitors of the MRLs) favoured "task statements in general terms, with maximum opportunity for creation and discovery". In spite of the temporary setback of the Mansfield Amendment, the United States has by and large been able to keep these ideals. Certainly, it seems that Senator Mansfield has had a less baleful influence than Lord Rothschild in Britain.

For non-American readers, there are some *longueurs* in the book because it concentrates exclusively on the United States and its achievements. But the diet is by no means an unvaried one of self-congratulation. DiSalvo bewails the neglect of the creation (synthesis) of novel bulk materials and compares it with the French penchant, orchestrated by the Centre National de la Recherche Scientifique, for combining synthesis of novel materials with immediate assessment of properties, often in the same laboratory. This is one of several places in the book where the need is pressed for materials scientists to give more attention to solid-state chemistry.

This volume represents a landmark in the literature of materials science and engineering. It deserves a wide readership, not only among materials practitioners but also among specialists in science policy. □

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

Life of the running waters

J. Green

The Ecology of River Systems. Edited by B.R. Davies and K.F. Walker. *Dr W. Junk: 1986. Pp.793. Dfl. 350, \$148, £97.*

THE study of river ecology came of age only 17 years ago with the publication of Noel Hynes's *The Ecology of Running Waters*. It is most appropriate, then, that Hynes has written the foreword to the present book. Together with the introduction by the editors, he provides the integrating part of the book, distilling into a few pages the modern approach to the subject. There has been a gradual realization that a river system needs to be seen in relation to its entire catchment area — lakes, reservoirs and floodplains are essential components of any study. The emphasis throughout the book is therefore rightly on the drainage basin as a whole, and where dams have been built they are given their fair share of attention.

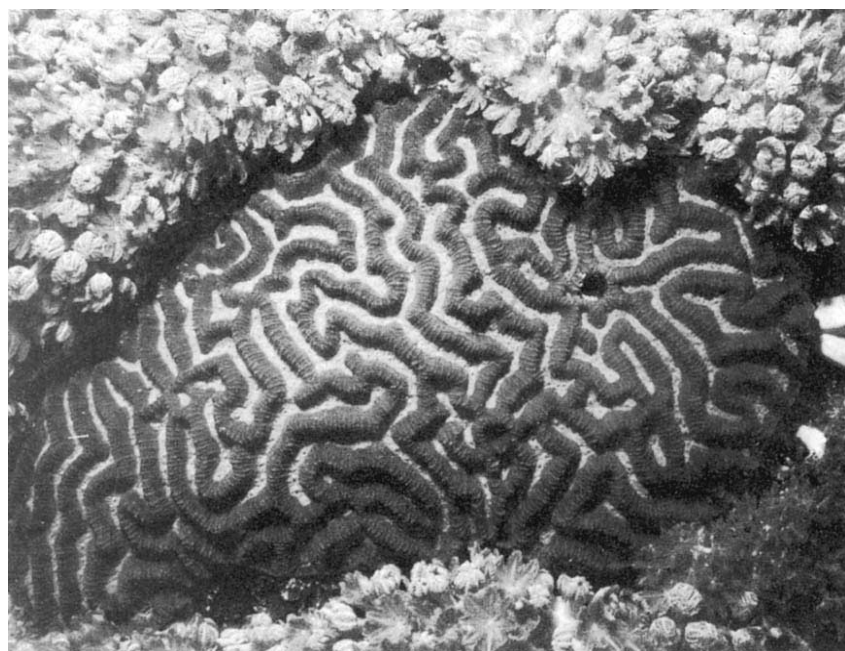
The main part of the book deals with the ecology of 13 major rivers outside Europe. Africa is best covered, with six, the Americas are next with five, while Australia and Asia have but one each. A huge amount of information is included, much of it not readily available elsewhere, and certainly not in the readable form that has been achieved by most of the authors.

The great value of these accounts is that they enable ecologists to make their own

comparisons, between the systems described here or with some other river system on which they may be working. The gaps in each of the contributions reflect the current state of our knowledge, and by clarifying just how much, or how little, we know, the individual chapters should stimulate further research. One of the points to emerge is that for nearly half of these large rivers we do not have basic information on all the main components such as the aquatic plants, plankton, benthos and fishes. Filling these lacunae in knowledge will not be an easy matter; sampling benthos, for example, requires boats, heavy equipment and considerable manpower for sorting, and few governments are willing to give priority to such work.

From the biogeographical point of view, we require similar accounts of yet other rivers. This is not an advocacy of a purely empirical approach. For our understanding of river systems to develop to a stage where generalizations and predictions may be possible, we need basic comparative data on rivers from a wide range of climatic and geological conditions. Is it too much to hope that in ten years' time we may get another book dealing, say, with the Indus, Ganges, Irrawaddy, Yangtze, Amur, Lena, Yenisei and Ob? That would form a fascinating series, and if the Mississippi and Orinoco were added we would then have, combined with the present book, the foundations for a fully scientific appreciation of some of the most vital and misused parts of our planet. □

J. Green is a Professor in the School of Biological Sciences, Queen Mary College, Mile End Road, London E1 4NS, UK.



Labyrinthine life — *Platygyra daedalea*, a brain coral which forms colonies a metre or more in diameter. The picture is a black-and-white reproduction of a colour photograph in *Red Sea Invertebrates* by Peter Vine, a large-format, beautifully illustrated book recently published by IMMEL, Ely House, 37 Dover Street, London W1X 3RB, UK. Price is £39.50.

Interacting ideas

Michael Rugg

The Brain Code: Mechanisms of Information Transfer and the Role of the Corpus Callosum. By Norman D. Cook. Methuen: 1986. Pp.256. £30, \$35.

THE discovery of a single principle which brought order to the apparently horrendous complexities of the workings of the brain would, to put it mildly, be very welcome. The "brain code", according to Norman Cook, is such a principle, and is exemplified by the way the two cerebral hemispheres interact via their largest interconnecting fibre tract, the corpus callosum.

Racing through such topics as comparative anatomy, the evolution of intelligence, and the nature of sensory representation and processing in the cortex, Cook derives and enlarges upon his "inhibitory topographical mapping" theory of inter-hemispheric communication. This theory is based on the twin assumptions that the corpus callosum interconnects homotopic cortical columns, and that the fundamental mode of action of this structure (with the exception of fibres interconnecting sensory cortices) is inhibitory. Activation of a cortical column in, say, the left hemisphere gives rise to inhibition of a homotopic column on the right side, but to excitation of immediately surrounding columns because of release from "surround inhibition". The upshot is that for any pattern of cortical activation on one side, a complementary pattern (likened to a photographic negative) will be observed on the other. This principle of cortico-cortico communication is the "brain code", and is considered by Cook as likely to be as fundamental to neuropsychology as the interaction between DNA and RNA is to biology.

Despite its scope, numerous high-quality illustrations and commendably clear style, this is not a textbook; areas of controversy are frequently discussed with virtually no reference to original work, and there are many minor errors of fact and occasional howlers (as in the statement on p.162 that Wernicke's area is located in parietal, rather than temporal, cortex). The book therefore stands or falls on its theoretical contribution. This, unfortunately, stems from a highly selective review of the literature on hemisphere specialization, and a naive conceptualization of how psychological functions are represented in the brain. For example, understanding a word is assumed to occur via excitation of a single cortical column corresponding to the word's denotative meaning in the left hemisphere, and, in the right hemisphere, excitation of a population of columns surrounding the

inhibited homotopic column to provide the connotative meaning. This theory is at variance both with observations of the way word meanings are lost following brain damage, and with the currently popular idea that at least some cortical areas act as parallel distributed systems in which information is represented in a spatially diffuse manner.

Cook quite correctly points to the importance of explaining how the two hemispheres interact to produce a single functional system. While criticizing others for falling back on simplistic dichotomies to characterize the different psychological functions of the two hemispheres, however, he proposes one himself, using the metaphor of an "executive" left hemisphere acting under the supervision of a "board of directors" in the right hemisphere.

Analysing the past

R.E.M. Hedges

Current Scientific Techniques in Archaeology. By P.A. Parkes. Croom Helm, London/St Martin's Press, New York:1986. Pp.271. £30, \$35.

THE HISTORY of the application of scientific techniques to archaeological questions is a fascinating one, and starts at least as far back as Sir Humphry Davy's analysis of the mysterious pigment 'Egyptian Blue'. Ancient authorities, from Pliny on, pronounced upon its supposed composition, and recent publications from the British Museum Research Laboratory on the same material show that Sir Humphry has not had the last word.

In one sense, scientific analysis does no more than extend the judgements we all make, as between stone and metal artefacts, for example. It then falls to archaeologists to divide these objects into categories, for instance of Stone or Iron Age, or, to be more up to date, Hunters or Pastoralists. But the methods concerned have become ever-more sophisticated, while the information provided is increasingly subtle. Such information might be genetic, environmental, technological, temporal, spatial; it might not relate readily to the framework of existing archaeological thought. And the methodology itself is liable to remain a mystery to all but the scientifically initiated.

It is the latter issue, the problem inherent in the widespread use of results of advanced scientific techniques by archaeologists, that this book addresses. The perspective has been narrowed down to consideration of most of the techniques within the physical sciences that are found to be helpful at present. It will therefore be primarily of interest to those archaeologists who want to 'bone up' on

sphere. He further argues that the respective abilities of each hemisphere are an important factor in determining personality differences between individuals, and are amenable to psychometric measurement. Such ideas have been the subject of trenchant criticism, on both conceptual and methodological grounds, from J.G. Beaumont and colleagues (*Cog. Neuropsychol.* 1, 191 - 212; 1984).

It seems doubtful that *The Brain Code* will achieve its highly ambitious goal. But its directness and clarity highlight our ignorance of many aspects of brain function, and will force readers to examine their own assumptions about the subject. □

Michael Rugg is a Medical Research Council Senior Research Fellow in the Department of Psychology, University of St Andrews, St Andrews KY16 9JU, UK.

the range of modern methods available. The book is organized by individual techniques — some 30 of them are included — and its strength lies in this and in the generous collection of recent applications which illustrate the various methods.

Although I regret the absence of a global view, the discussion of each method is for the most part accurate and sensible; I found only a few points to quibble at in those fields I know best, and apart from the caveat that much of the recent literature is cited rather too uncritically, I would recommend the book for its main purpose. It also makes more attempt than most to set down simple practical knowledge for the archaeologist, for example in how to provide samples, or in its appreciation of such experimental issues as the statistical presentation of results. The last chapter on computers may help the really frightened not to run away. But although a wide range of topics is considered, the balance seems to have been determined more by the author's research interests than by the relative power of different techniques. Archaeomagnetism is treated in 38 pages (the largest section), while, for example, oxygen isotope stratigraphy receives less than three and the use of stable isotopes in the study of palaeo-diets is covered in under six.

What the book lacks above all, however, is any sense of intellectual excitement. It conveys little of the ingenuity of the scientific thought involved, or of the changing perspective of archaeological scholarship. This lack of sparkle is intensified by the dull and even tone of the writing, which I fear might widen rather than bridge the gap between the two cultures. Finally, readers must surely expect better standards of production for a book of this very considerable price. □

R.E.M. Hedges is Director of the Radiocarbon Accelerator Unit at the Research Laboratory for Archaeology, University of Oxford, 6 Keble Road, Oxford OX1 3QJ, UK.

Automated high-speed DNA sequencing

Akiyoshi Wada

The Japanese project for the development of a super-sequencer for DNA analysis has reached the point at which estimates of cost can be attempted.

THE techniques developed in the past decade for determining the base sequence of pieces of DNA have influenced the goals of molecular biology in at least two important ways. First, they have made possible a deeper exploration of the molecular mechanisms of life, as in many studies of the regulation of the activity of genes and in the localization of the genes responsible for various cryptic genetic diseases. But, second, DNA base sequencing yields data which are of interest in their own right, both in comparisons of genes with similar functions from different organisms and in the compilation, at the highest possible resolution, of the genetic maps of individual organisms. DNA base sequencing is bound to be an important part of the activity of molecular biologists for centuries to come.

As information accumulates from the sequencing of DNA from various organisms, it will become an ever more valuable asset for society as a whole, with potentially useful applications as different as a better understanding of cancer and the identification of the causes of various genetic defects and traits¹. Unfortunately, the amount of information that will eventually have to be analysed is several orders of magnitude greater than the present rate of production of base-sequence information. Sequence data are at present being added to the databanks at the rate of 10^6 bases a year. But the human genome alone consists of 3×10^9 base pairs partitioned between 23 pairs of chromosomes, and the magnitude of the task ahead dwarfs the capabilities of the present methods.

Automation

The history of human technology is that of the replacement of repetitive sequences of individually complex tasks by automatic machines. The processes of DNA base sequence analysis are essentially of this kind, and are therefore prime candidates

for automation, especially now that there is general confidence in the reliability of the results of these procedures.

The potential benefits of the automation of DNA base sequencing include the relief of skilled researchers from repetitively mechanical and often tedious manipulations (which is not to say that the importance of human judgement in the interpretation of data should be undervalued) and the more rapid accumulation of data. But automation would be justifiable only if there is a substantial increase the speed of data accumulation and a reduction in the cost of analysis.

At present, one skilled researcher may, in favourable circumstances, be responsible for the sequence of up to 1,000 bases a day at a cost of something like \$1 per base. It would seem that the output of an automatic DNA base sequence analyser should be some orders of magnitude greater, perhaps 1 million bases (1Mb) per

machines that could be the successive elements of a mass-production line for sequencing DNA. We have now reached the point in the planning at which all the elements of a mass production line exist, together with the interfaces between them.

In our choice of particular processes for carrying out successive chemical steps in the sequencing of DNA, we have been guided by the principles that apply in the design of engineering production lines. In particular, our choice of successive processing elements has been determined by the need for processing times at successive stages to be well matched². This, for example, has led us to the use of the Sanger procedure which is much faster than the Maxam-Gilbert method for base-specific fragmentation of DNA; even so the Maxam-Gilbert method will still be important as a relief in our system for the cases for which the former method is inadequate.

We must emphasize that our goal is not that of a fully robotic system, but of a DNA sequencing line capable of yielding the sequence of 1×10^6 bases (1 Mb) per day in which skilled human operators would play a crucial part in monitoring the accuracy of the results. Although the conceptual design of our DNA sequencing line is obviously relevant to present interest in the sequencing of the human genome, we believe that there is already a sufficient need for a sequencing service to keep a system such as ours fully occupied, perhaps on an international basis.

In any event, to match the high-speed processing of such a super-sequencer, it will be necessary to establish an international DNA-sequencing centre, and to open an international channel of communication from the receipt of researchers' samples by the centre to the reporting of results of the analyses to those researchers. If such a channel is convenient and economically interesting to users, many samples will be received, increasing the number of steady users and, in turn, increasing the efficiency of analysis.

Development

A sequencing machine using the Maxam-Gilbert protocol^{3,4} is now commercially available from Seiko; it can handle up to twelve samples simultaneously. For each sample, the output from this machine is a

Table 1 Estimated costs per nucleotide base, 1 Mb/day sequencer

	Cost (yen per base)
1. Sample handling, data reporting*	0.1
2. Scission, cloning, plaque selection†	0.04
3. Purification (Y500/250 bases for filter/resin)	2.0
4. Dideoxy-sequencing (kits Y100,000/100 samples)	4.0
5. Fuji Gensor gel film‡	2.0
6. Gel electrophoresis	≈0.0
7. X-ray film, processing (Y500/film)	0.5
8. Gel reading	≈0.0
9. Instrument depreciation§	0.3
10. Personnel costs¶	0.24
Total	Y9.18
For shotgun method (× 3)	Y27.5
(In US dollars = \$0.17)	

Because the cost estimates of materials in items 3, 4 and 5, which represent a large portion of the total, are based on present retail prices, bulk purchase for use in a DNA sequencing factory could bring a 50 per cent reduction of the total cost, yielding 10 cents per nucleotide base.

*Assuming handling, mailing costs for one 40 kb sample of Y4,000. †M13 cloning of ≈ 500 base sequences, selection of 500 plaques yielding ≈ 250-base sequences. ‡Y2,500 per film, each giving 1,250 bases. §Assuming an initial cost of Y300 million depreciated over five years and operation of the sequencer for 200 days per year. ¶Assuming Y8,000 per day for each of 30 persons.

day². Moreover, the cost of analysis should be reduced by an order of magnitude by virtue of mass-sequencing (Table 1).

These are some of the reasons why, five years ago, the Science and Technology Council of Japan formed a committee, of which I am chairman, to study the opportunities for automation in processing DNA sequences. We have been concerned, in collaboration with industrial companies on the one hand and molecular biologists on the other, to develop the

set of four groups of DNA fragments (which differ by terminating at the bases G, G or A, T or C and C respectively), which may then be subjected to electrophoresis. Unfortunately, the cycle time for this process is 12 hours, corresponding to an output of 6,000 bases a day or one base every 14.4 s, rather less than the production rate of 1 Mb a day (0.086 s per base) which is our goal.

In collaboration with Seiko, we are therefore developing a machine for automatic analysis of DNA by the Sanger method. This total system will include robotic machines for plaque selection and for extraction and purification of DNA from cosmid or shotgun libraries. From experience with a prototype constructed by Seiko under the guidance of E. Soeda of the Institute of Physical and Chemical Research at Wako, Saitama, we estimate the daily output of such a system at 300,000 bases a day, or one base every 0.28 s.

Experience shows that the resolution of the gel separations derived from the output of the Sanger machine compare favourably with those of films derived by the manual manipulation of the DNA samples. We are sure that a machine of this kind can be adapted for continuous operation, with DNA sub-samples, 250 bases or so long, supplied to the machine in accurately indexed wells in a virtually endless plastic belt. A relatively modest improvement of the capacity of this machine would achieve 1 Mb a day.

Of the key components of an automatic sequencing line, that for the electrophoretic separation of DNA fragments is the most taxing. In molecular biology laboratories, acrylamide gel films for electrophoresis are customarily prepared manually; the procedure usually requires that, after electrophoresis for between 12 and 24 hours, films should be dried and then exposed to X-ray films (where the detection of DNA fragments is based on the use of radioactively labelled bases).

In our project, Fuji Photo Film Co. of Tokyo has developed a machine for the manufacture of prepackaged acrylamide-gel films, each of which is 20 cm $\times$ 40 cm and 0.2 mm thick, with 20 pre-cast slots for sample injection at the top. Each film is covered with acetate film front (0.06 mm) and back (0.18 mm). The front covering is thin enough for the films to be read directly by exposure to X-ray film, thus obviating the need for drying. The product is now commercially available under the name of Gensor gel at a cost of US\$16 per film. We hope that, by the production of films with a controlled density gradient along the direction of electrophoresis (which is within the capability of Fuji's machine), it will be possible to improve substantially on the precision with which DNA fragments may be distinguished from each other.

Fuji's pilot machine produces films at the rate of 1,000 units a day. If each film is used for the electrophoresis of five DNA sub-samples of 250 bases each in $4 \times 5 = 20$ lanes, this output is already sufficient for our goal of a production rate of 1 Mb per day. Furthermore, well-characterized and standardized films made by such a precisely controlled mass-production system will help guarantee the precision of the automation system.

We are engaged on the development of a machine in which the output of samples from a Sanger sequencer can be subjected to electrophoresis under standardized controlled conditions for periods in excess of 12 hours. Of necessity, such a machine will contain several hundred films at once. We anticipate no exceptional problems with the mechanical handling of either the gels or the X-ray films which are the ultimate products of this part of the system.

Alternatives

At least for the time being, we have based our development on the use of the conventional techniques of electrophoresis, and in particular on the use of radioactively labelled bases followed by autoradiography, but this is not to imply that we underestimate the recent development of other methods for reading and interpreting the information embodied in DNA fragments. Rather, we consider that the conventional methods are well tested, quite appropriate for parallel processing, capable of high sensitivity and accuracy and, at the same time, relatively cheap. We acknowledge that, with further development, techniques such as specific labelling of the four nucleotide bases with characteristic fluorescent dyes may be advantageous in an automated system, if their cost-performance becomes attractive.

Meanwhile, we intend that the interpretation of the autoradiography films from our system should be based on one of the automatic scanners developed independently by Hitachi Software and by Seiko. These scanners use charge-coupled device (CCD) array cameras or contact type CCD array sensors for digitizing electrophoresis signal bands. Correct recognition of the true signal on the gel, the necessary correction of deformed band shapes and the accurate elimination of spurious bands are handled by computer software.

With an autoradiograph of average quality, the error in reading is less than one per cent up to 250 bases and the maximum reading rate is 1.4 s per base $\approx$ 60 kb per day. Although the performance, both in reading speed and accuracy, may be a little behind that of an alert researcher, the machine can work 24 hours a day with constant precision, without training or complaining. Improvement in speed and accuracy by a factor of 10 may be attainable by using a better optical processor

and a computer of higher grade. A checking system analysing the two identical (or complementary) strands by parallel analysis lines, which is the virtue of automation, may be introduced, by which the error level is reduced to the order of 10^{-4} ; only patterns unidentifiable by machine are given to an expert for judgement.

Opportunity

Whether one likes it or not, we are now entering an age of mass-reading of DNA base sequences. The need for a high-performance sequencing machine is clear, and the time is ripe for its development because of progress in related fields: the accelerated speed of computation, increasingly widespread appreciation of the value of a DNA database, the development of highly sophisticated measuring devices and robotic technology and the growth of chromosome and gene libraries, especially those of humans. The advantage of a well-organized mass-analysis line is obvious, as already shown by many mass-production lines in industry; it is not only the high speed of processing and the economy of labour and cost, but also the ease of quality control and the reproducibility which ensure confidence in the result.

The significance of the large DNA sequencing centre or 'factory' should be measured not only by the current demand for such a super-machine, but also by the attractiveness of such a centre to individuals and laboratories worldwide. Such a centre would be the recipient of many samples, thus increasing the number of its regular customers and, in turn, the efficiency of its analysis. The arrangement would be self stimulating.

In the twenty-first century, we foresee that DNA-sequencing super-centres will be set up in several countries and will become symbols of the effort of nations to broaden and build on human knowledge, taking their place beside other existing symbols such as large particle accelerators, giant telescopes and far-reaching programmes of space research.

This article has emphasized the Japanese super-system; readers interested in the current status of the automation of various units in DNA sequencing are referred to the reviews by Martin and Davies⁵ and by the present author², for more detailed discussion of equipment and for more complete references. $\square$

Akiyoshi Wada is Professor of Biophysics, Physics Department, University of Tokyo, Japan.

1. Dulbecco, R. *Science* **231**, 1055-1056 (1986).
2. Wada, A. & Soeda, E. *Proc. 4th Congr. Fed. Asian Oceanian Biochemists* (Cambridge University Press, in the press).
3. Wada, A., Yamamoto, M. & Soeda, E. *E. Rev. sci. Instr.* **54**, 1569-1572 (1983).
4. Wada, A. *Nature* **307**, 193 (1984).
5. Martin, W. J. & Davies, R. W. *Bio/Technology* **4**, 890-895 (1986).

Strategies for antiviral therapy in AIDS

Hiroaki Mitsuya & Samuel Broder

The Clinical Oncology Program, Division of Cancer Treatment, National Cancer Institute, Bethesda, Maryland 20892, USA

The retrovirus that causes AIDS has revealed enough of its life history for a variety of therapeutic strategies to be apparent. Some of these are suitable for immediate application in clinical trials or have already yielded positive results in some patients.

ACQUIRED immune deficiency syndrome (AIDS) was initially defined as the development of either an opportunistic infection or Kaposi's sarcoma (an unusual neoplasm that had previously been recognized to be associated with certain immunosuppressed states) in a person without a known cause for immunodeficiency^{1,2}. It soon became apparent that these patients had a cellular immunodeficiency characterized by an inexorable depletion of helper/inducer (T4⁺ or CD4⁺) T cells, and within three years it had been proven that a retrovirus, which has the capacity to replicate within cells of the immune system and so lead to a profound destruction of T4⁺ T cells, is the cause of AIDS^{3,4}. The retrovirus is the third known human T-lymphotropic virus (HTLV-III). The virus has also been called lymphadenopathy-associated virus (LAV)⁵ or AIDS-associated retrovirus (ARV)⁶ and it has more recently been proposed that it is designated as human immunodeficiency virus (HIV)⁷.

The immunodeficiency of AIDS is manifested as a loss of delayed-type cutaneous hypersensitivity reactions, a loss of certain *in vitro* T-cell proliferative responses, excessive immunoglobulin production by B cells, perhaps due to abnormal regulation coupled with a T-cell dependent mitogenic effect mediated by the virus, and a loss of *in vitro* cytotoxic T-cell responses. In addition, the opportunistic agents which most characteristically cause illness in HTLV-III infected patients (*Pneumocystis carinii*, *Mycobacterium avium* and *Candida albicans*) are normally controlled by cell-mediated (T-cell) immune responses. Once established, the immunodeficiency state is generally progressive and leads to death. The hope is that this progression can be handled and even reversed by antiviral agents that prevent replication of the virus and its cytopathogenicity in cells of the immune system.

In addition to its effects on the immune system, there is a growing recognition that HTLV-III can cause serious neurological diseases, ranging from peripheral neuropathy to fulminant dementia, with or without a clinically evident immuno-

deficiency⁸⁻¹⁰. Perhaps infected macrophages play a role in the traffic of virus across the blood-brain barrier^{11,12}. Although it is still not established precisely how the virus enters the brain and how it brings about neurobiological damage, it is a virtual certainty that successful therapeutic strategies must address the consequences of, and probably prevent, viral replication in the central nervous system. Fortunately, this goal may not be out of reach for certain therapies now under study; we will return to this later.

As with any virus, the different stages in the life cycle of HTLV-III present a variety of distinct potential targets for antiviral agents. It is the purpose of this review to outline these targets, which are summarized in Table 1, and to focus on the reverse transcriptase inhibitors since there has already been some success in this area, notably with 3'-azido-3'-deoxythymidine (AZT)^{13,14}. The testing of new antiretroviral agents has been facilitated by the availability of rapid and sensitive *in vitro* screening systems that determine whether a putative drug can inhibit the replication and T-cell killing activity of HTLV-III¹⁵. At the outset, it is worth stressing that effective therapy of HTLV-III infection may well depend on a combination of strategies that attack multiple steps in viral replication and cytopathogenicity, in part because the emergence of drug-resistant strains might be less likely.

HTLV-III as a retrovirus

In the era before the aetiology of AIDS was recognized, one of the most notable features of retroviruses was their capacity to induce neoplastic transformation in infected target cells; hence the expressions 'RNA tumour virus' and 'leukaemia virus' appeared in the literature to denote this category of virus¹⁶. But the virus that causes AIDS has not been shown to have a transforming capacity. (It is only linked to the causation of cancer through its immunosuppressive activity.)

By definition, retroviruses replicate through a DNA intermediate (that is, at one step of their cycle of replication, genetic information flows from RNA to DNA, a reverse or 'retro' direction^{17,18}). Reverse transcriptase, the viral DNA polymerase that catalyses this step, is encoded by the *pol* gene of the virus, a gene that is conserved to a considerable extent in its amino-acid and nucleotide sequences among retroviruses¹⁹. These viruses have four major structural components: a core of genomic RNA; group-specific antigen (*gag*) proteins, which are involved both in the structure of the core and assembly of the virion; a lipid bilayer; and an outside envelope glycoprotein. HTLV-III is the most complex retrovirus yet characterized^{20,21}, in that it contains at least four genes that are not found in previously defined retroviruses (Fig. 1). The functions of two of these four genes, *src* and 3'-*orf*, are not known. The functions of the other two genes, *tat*-III^{22,23} and *art*²⁴ or *trs*²⁵, will be discussed later.

Cell binding and entry

The first step in the infection of a cell by HTLV-III is its binding to the target cell receptor. In the case of helper/inducer T cells, this receptor seems to be linked to the cell-surface protein that is recognized by T4 or CD-4 antibodies^{26,27}. The process of

Table 1 Stages in the replicative cycle of a pathogenic human retrovirus which may be targets for therapeutic intervention

Stage	Potential intervention
Binding to target cell	Antibodies to the virus or cell receptor
Early entry into target cell	Drugs that block fusion or interfere with retroviral uncoating
Transcription of RNA to DNA by reverse transcriptase	Reverse transcriptase inhibitors
Degradation of viral RNA in an RNA-DNA hybrid	Inhibitors of RNase H activity
Integration of DNA into host genome	Drugs which inhibit <i>pol</i> gene-mediated 'integrase' function
Expression of viral genes	'Anti-sense' constructs; inhibitors of the <i>tat</i> -III protein or <i>art</i> / <i>trs</i> protein
Viral component production and assembly	Myristylation, glycosylation and protease inhibitors or modifiers
Budding of virus	Interferons

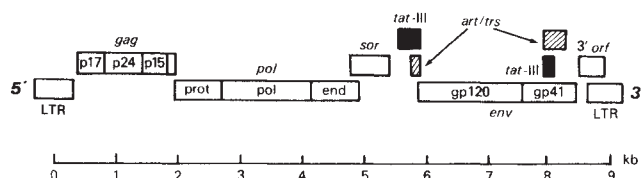


Fig. 1 Organization of the HTLV-III (human immunodeficiency virus) genome. The virus is known to contain at least seven genes, flanked at each end by long terminal repeat (LTR) segments, which contain promoter and enhancer sequences for viral transcription, as well as recognition sites for viral endonuclease (integrase) activity. Adapted from refs 20–25.

specific binding between the T4 protein and parts of the viral envelope glycoprotein may be vulnerable to attack by antibodies either to the virus or to the receptor, and theoretically certain chemicals or small peptides²⁸ could be designed to occupy the receptor and accomplish the same thing.

It might be worthwhile deviating from the chronology of the retroviral life cycle briefly to address the cytopathogenicity of the virus that causes AIDS. From one perspective, the T4 protein and the envelope glycoprotein of the virus determine events both at the beginning and end of the life cycle of HTLV-III within susceptible T cells. In addition to a role in the receptor-mediated entry of the virus into target T cells, the T4 protein plays a part in the susceptibility of a target T cell, once it begins to produce virus, to killing by that virus. Precisely how HTLV-III destroys T cells *in vivo* is not known. The cytopathic effect is thought to be mediated in part by an interaction between the T4 protein and the HTLV-III envelope protein that brings about lethal cell-to-cell fusions (syncytia) or a surface autofusion phenomenon that destroys the integrity of the cell membrane^{29,30}. But other factors are also thought to play a role in the cytopathic effects of the virus. Recent results raise the possibility that the carboxyl terminus of the envelope protein, a region different from that which directly binds to the T4 protein, is important in the destruction of T cells by the virus³¹.

Conceivably, therapeutic agents could be designed to alter the properties (for example, the lipid composition) of the surface of the virus or target cell so as to reduce the infectivity or cytopathic effects of the virus. An alternative target is the envelope protein itself. Although there is considerable variation in the protein from one viral isolate to another, the range of alterations in the binding site is most probably constrained by the need to bind to T4, which is relatively constant in structure. An antibody directed against this site might bind to (and neutralize) most strains of HTLV-III and perhaps kill infected cells as they begin to express envelope antigens, so that spread of virus to uninfected cells can be reduced.

Thus, monoclonal antibodies to the envelope protein could have a therapeutic role in patients with AIDS or related diseases. We have recently been able to produce a complement-fixing human monoclonal antibody to the major envelope glycoprotein of the first known human T-lymphotropic retrovirus (HTLV-I)³², and the same approach could be used to develop similar human monoclonal antibodies to HTLV-III. A potential difficulty of this approach, however, is that virally infected cells might make infectious cell-to-cell contacts. Antibodies might not gain access to relevant epitopes of the envelope protein under such circumstances. Also, it has been shown that AIDS can occur in the face of what *in vitro* seem to be neutralizing antibodies to HTLV-III. Whether that is because the titres of such antibodies are low or because such antibodies do not block epitopes that mediate *in vivo* cytopathogenicity is under investigation.

After binding to a cell, HTLV-III enters the target cell by an incompletely defined mechanism, most probably a fusion process. It is conceivable that drugs could be developed to block this step, just as calmodulin antagonists block the entry of

Epstein-Barr virus into B cells. Another theoretical target is the stage of 'uncoating' of the virus after it enters a target cell. In this stage, the virus loses its coating of envelope protein and its RNA is released into the cytoplasm (each virion is thought to convey a dimer of two identical genomic RNA subunits into the cell). Pharmacological agents that block viral uncoating might eventually be developed for the treatment of AIDS.

Reverse transcription

Uncoated viral RNA is used as a template for the production of DNA by reverse transcriptase. This enzyme has become a prime target for antiretroviral agents both because it should be possible to find inhibitors that will discriminate between this enzyme and the DNA polymerases of the host cell, thus avoiding side effects, and because a great deal is already known about reverse transcriptase. HTLV-III reverse transcriptase uses a lysine transfer RNA as a primer to make a minus strand DNA copy of the viral RNA as an RNA-DNA hybrid²¹. Well as having reverse transcriptase activity, the retroviral DNA polymerase possesses an inherent RNase H activity that specifically degrades the RNA of the RNA-DNA hybrid and, in theory, inhibition of this process would suppress viral replication. The reverse transcriptase then catalyses the production of a positive strand DNA, and eventually, a double-stranded DNA is formed.

The reverse transcriptase of HTLV-III has been purified and seems to exist as a p51 and a p66 molecule³³. Large quantities of the HTLV-III reverse transcriptase should become readily available because the enzyme has been expressed in bacteria³⁴ and yeast (P. J. Barr and P. A. Luciw, personal communication) using recombinant DNA technology. This should aid the testing of potential reverse transcriptase inhibitors. The use of such agents in patients with AIDS is predicted on the assumption that there is continuing viral replication in the disease and that its inhibition would permit some regeneration, or at least prevent further deterioration of the immune system. Two drugs (suramin and HPA-23), which were chosen early on for clinical trials because they inhibit reverse transcriptase, have not appeared to confer clinical benefits on patients^{35–37}. Nevertheless, agents which inhibit reverse transcriptase have the most likelihood of achieving an immediate clinical impact on the virus. Several promising agents that inhibit this enzyme are either already available, having been developed for the therapy of conventional viral diseases (for example, phosphonoformate³⁸) or are being developed specifically for AIDS on the basis of *in vitro* screening systems for activity against HTLV-III^{13,15}. We will return to one class of such inhibitors, the dideoxynucleosides, one member of which (AZT) has already been shown to prolong survival in patients with AIDS.

Integration latency and reactivation

The DNA copy of HTLV-III may be circularized during or soon after its formation, and apparently can either remain in an unintegrated form or can become integrated into the genome of the host cell. It is possible that chemicals could be developed to interfere with the viral endonuclease or 'integrase' (thought to be another function of one of the *pol* gene products) that mediates this integration step. Later, perhaps after activation of the infected cell, the DNA is transcribed to messenger RNA and viral genomic RNA by host cell RNA polymerases, and this RNA is then translated to form viral proteins, again using the biochemical apparatus of the host cell.

It has recently been shown that HTLV-III has a gene (*tat-III*) coding for a diffusible protein that markedly enhances the expression of other viral genes and viral replication. The *tat-III* gene, like the *tat* genes of the other two known human pathogenic retroviruses, is so called (by acronym) because it was originally thought to mediate a *trans*-activation of transcription—that is through a mechanism that affects the transcription of genes not in direct proximity to itself^{22,23}. But ideas about the *tat-III* gene are still evolving^{39,40}. The *tat-III* protein seems

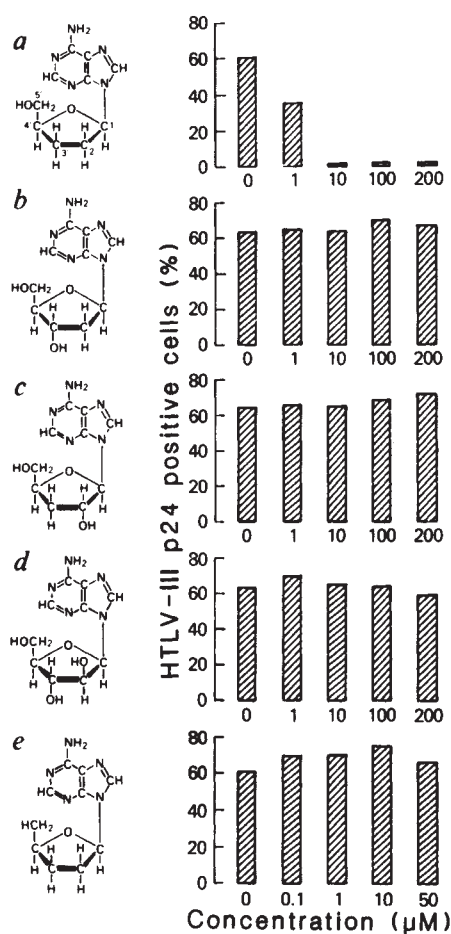


Fig. 2 Protection against infection by HTLV-III by adenosine congeners. H9 cells which are permissive for HTLV-III replication⁴ were used as target cells. Cells (10^5) were pre-exposed to hexadimethrine bromide (HDB), exposed to HTLV-III_B (3,000 virus particles per cell), and resuspended in culture tubes in the presence of various concentrations of adenosine congeners: a, 2',3'-dideoxyadenosine; b, 2'-deoxyadenosine; c, 3'-deoxyadenosine; d, 9- β -D-arabinofuranosyladenine; e, 2',3',5'-trideoxyadenosine. On day 9 of culture, the percentage of cells expressing p24 gag protein of HTLV-III was determined. H9 cells not exposed to virus were all negative for gag protein. Similar results are obtained with ATH8 cells⁴⁶.

to cause an increase of viral products at a transcriptional level. However, the protein may also increase viral expression by acting beyond transcription of DNA to RNA, by enhancing mRNA stability and/or translational efficiency. Whatever the mechanism, the *tat*-III protein is thought to provide the virus with a positive feedback loop by which a viral product can amplify the production of new virions. The protein is small (86 amino acids) with a cluster of positively charged amino acids, and is thought to affect the synthesis of other proteins by binding to critical regulatory sequences at the 5'-end of viral mRNA. It may be possible to find drugs that inhibit the *tat*-III product itself, a crucial nucleic acid binding site for this protein, or both.

Employing the same mRNA that codes for *tat*-III, but using a different reading frame, the *art* or *trs* gene of HTLV-III produces a different small (116 amino acids) positively charged protein, which is thought to function as a second essential *trans*-acting factor in viral replication^{24,25}. In the absence of this regulatory factor, *gag* and *env* encoded protein synthesis is severely diminished, so drugs that bind or inactivate this protein would be expected to inhibit viral replication. There is no doubt

about the existence of this new regulatory gene, but the mechanism of action of its product is controversial. The action was first described as an abrogation of negative regulatory effects on translation of viral mRNA that encodes HTLV-III structural proteins (hence the name *art* for anti-repression *trans*-activator)²⁴. It is possible that *tat*-III controls *art*, which in turn regulates the activity of *gag-pol* and *env*. More recently Feinberg *et al.*²⁵ have suggested that the gene product balances the amount of spliced and unspliced viral RNA to permit the preservation of *gag-pol* and *env* mRNA (hence the alternative name, *trs* for *trans*-acting regulator of splicing).

Although the precise mechanisms are matters of future study, it is clear that this retrovirus has evolved an astonishingly complex system of genetic regulation. Perhaps this is because of the race between viral replication within a cell and the destruction of the cell that the virus has commandeered, leaving no tolerance for inefficiency or improper timing in the synthesis of viral components. With luck, the very complexity of the virus could contribute to its defeat.

Protein production and assembly

It is conceivable that drugs that interfere with the structure and function of retroviral mRNA transcribed from integrated DNA in infected cells could be of therapeutic value in AIDS. One drug, ribavirin, is believed to act as a guanosine analogue that interferes with the 5'-capping of viral mRNA in other viral systems, and perhaps could be useful in retrovirally-induced disorders⁴¹. It is expected that the clinical efficacy of ribavirin in treating HTLV-III infection will be determined at the conclusion of studies now under way.

Another approach would be the use of 'anti-sense' oligodeoxynucleotides, already tested *in vitro*⁴². Basically this approach employs short sequences of DNA (or DNA that is chemically modified to enable better cell penetration and resistance to enzymatic degradation) whose base pairs are complementary to a vital segment of the viral genome. In theory such anti-sense oligodeoxynucleotides could block expression of the viral genome through a kind of hybridization arrest of translation or possibly through interference with the binding of a regulatory protein such as *tat*-III. Indeed, as the *tat*-III gene and the *art*/*trs* genes are expressed on the same mRNA transcript, it is possible that one anti-sense oligodeoxynucleotide could interrupt two functions that are vital for HTLV-III replication. In principle, it might be possible to achieve the same goal by constructing an 'anti-sense' virus (a retrovirus that has been genetically engineered to produce a stretch of mRNA that will bind to the messenger made by the wild-type virus).

The final stages in the replicative cycle of HTLV-III involve crucial secondary processing of certain viral proteins both by proteases (a function of one of the *pol* gene products) and by myristylating and glycosylating enzymes (provided by the host) as a prelude to assembly of infectious virions. Therefore, additional strategies for the treatment of AIDS might involve certain kinds of protease inhibitors or drugs which dampen or alter myristylation and glycosylation steps in the synthesis of viral components^{43,44}. Finally, retroviruses are released by a process of viral budding, which may be inhibited by interferons⁴⁵.

We have focused the discussion on how to suppress HTLV-III, but it might be worth noting that the virus could set off a chain of secondary events (such as autoimmune reactions, the production of toxic lymphokines) that are necessary for the expression of clinical disease. It is also intriguing to speculate that a combination of antiretroviral therapy coupled with bone marrow or lymphocyte replacement therapies might be successful in certain subsets of patients with HTLV-III infections.

Dideoxynucleosides

A broad family of 2',3'-dideoxynucleoside analogues can be metabolized to become potent chain-terminating inhibitors of HTLV-III reverse transcriptase. *In vitro* these analogues can

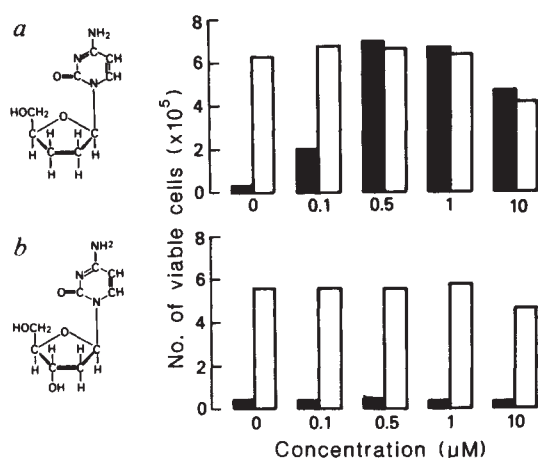


Fig. 3 Inhibition of the cytopathic effect of HTLV-III by 2',3'-dideoxycytidine. ATH8 cells¹³ (2×10^5), which are sensitive to the cytopathic effect of HTLV-III, were pre-exposed to HDB, exposed to a high multiplicity of infectious HTLV-III_B (2,000 virus particles per cell) in culture tubes (solid columns) in the presence of various concentrations of 2',3'-dideoxycytidine (a) or the normal nucleoside, 2'-deoxycytidine (b). Control cells (open columns) were similarly treated, but were not exposed to the virus. On day 5, total viable cells were counted. Under conditions of low multiplicity of infection, concentrations of drug ≥ 10 nM suppressed the virus.

completely inhibit the replication of HTLV-III and its capacity to destroy T-cell cultures at concentrations that are 10–20 fold lower than those that impair the proliferation and survival of target cells⁴⁶. Pioneering studies of several of these compounds have been accomplished over the past twenty years or so^{47–53} and, in triphosphate form, they are familiar to molecular biologists as reagents for the Sanger DNA-sequencing procedure⁵³. But their application to human antiretroviral therapy demands a new perspective. The dideoxynucleoside analogues are of special interest because they prove that a simple chemical modification of the sugar moiety can predictably convert a normal substrate for nucleic acid synthesis into a compound with a potent capacity to inhibit the replication and cytopathic effect of HTLV-III, at least *in vitro*⁴⁶.

Certain relationships between the structure and activity of adenosine analogues are explored in Fig. 2, which summarizes the capacity of five congeners of adenosine to act as inhibitors of HTLV-III. In these experiments, the adenine portion of the nucleoside analogue is kept constant whereas several simple changes in the sugar moiety are tested as variables. The reference agent is 2'-deoxyadenosine, which is a normal building block for DNA and so, as expected, has no antiretroviral activity in this system (Fig. 2b). But a simple reduction (removal of the oxygen) at the 3'-carbon of the sugar, creates a potent inhibitor of the retrovirus, 2',3'-dideoxyadenosine (Fig. 2a). A further reduction at the 5'-carbon, creating 2',3',5'-trideoxyadenosine (Fig. 2e) nullifies the antiretroviral effect, and neither 3'-deoxyadenosine (cordycepin) nor the *arabinosyl* derivative (Ara-A) inhibits the virus. It is important to note that these experiments were performed under conditions of high estimated multiplicity of infection¹⁵.

We are currently developing 2',3'-dideoxycytidine as a possible agent for treating patients infected with HTLV-III. On a molar basis, dideoxycytidine is the most potent antiviral nucleoside that we have tested. Under certain *in vitro* conditions, the drug can suppress HTLV-III at concentrations of 10 nanomolar or below. This dideoxynucleoside analogue was selected for clinical development because of its *in vitro* potency against HTLV-III⁴⁶ (see Fig. 3), relative resistance to cytidine deaminase (a major catabolic enzyme for cytidine analogues)⁵⁴, good oral bioavailability, straightforward pharmacokinetic clearance by

Terminated DNA Chain

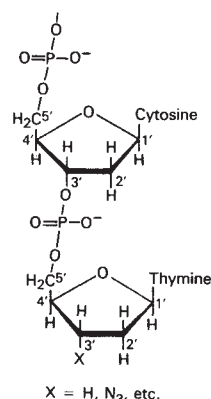


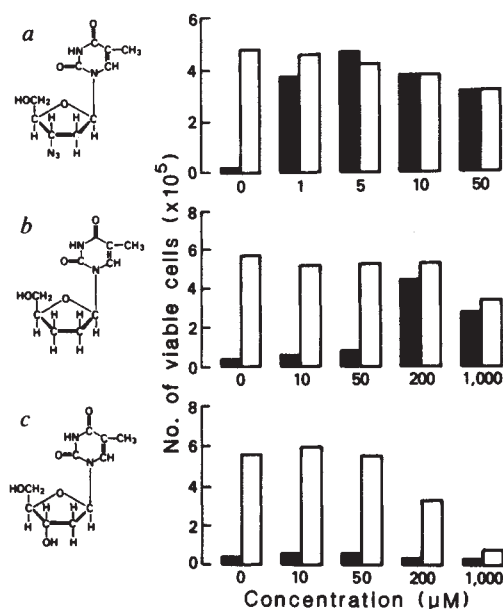
Fig. 4 Possible mechanism of activity against HTLV-III of 2',3'-dideoxynucleosides as triphosphate products. When the 3'-carbon of the deoxyribose is modified by certain substitutions (shown by X) that replace the normal 3'-OH, it is not possible to form 5' → 3' phosphodiester linkages that are necessary for DNA elongation in the replication of the virus from an RNA form to a DNA form. Similar considerations apply when the sugar is in a 2',3'-dideoxy (double bond between the 2' and 3' carbons) configuration.

the kidney, failure to reduce normal intracellular pyrimidine pools and comparative lack of toxicity when administered to animals. We have recently begun to administer this drug to patients.

Several issues related to the anti-retroviral effects of 2',3'-dideoxynucleosides are as yet unresolved, but it seems that as they are successively phosphorylated in the cytoplasm of a target cell to yield 2',3'-dideoxynucleoside-5'-triphosphates, they become analogues of the 2'-deoxynucleoside-5'-triphosphates that are the natural substrates for cellular DNA polymerases and reverse transcriptase. (It is generally thought that nucleoside-5'-triphosphates do not cross cell membranes and are not active as drugs because of their ionic character and comparatively low lipophilicity.) Such analogues could compete with the binding of normal nucleotides to DNA polymerases, or could be incorporated into DNA and bring about DNA chain termination because normal 5' → 3' phosphodiester linkages cannot be completed (see Fig. 4). At concentrations that are achievable in human cells, a dideoxynucleotide analogue can serve as a substrate for the HTLV-III reverse transcriptase, resulting in DNA chain termination after the addition of a single dideoxynucleotide residue⁵⁵.

Pyrimidine and purine dideoxynucleoside analogues can inhibit the *in vitro* replication and pathogenic effects of a range of animal and human retroviruses, even when the pathogenic effect being monitored (transformation) requires only a single round of replication; moreover, with certain lentiviruses (a family of animal retroviruses related to HTLV-III⁵⁶), these drugs can reduce the *in vitro* viral infectivity by more than five orders of magnitude⁵⁷. We have found that two dideoxynucleosides, 2',3'-dideoxycytidine and AZT(3'-azido-3'-deoxythymidine) can block the infectivity of HTLV-I against helper/inducer T cells *in vitro* (S. Matsushita, H.M., M. S. Reitz and S.B., unpublished data). Indeed, so far we have not observed any type or strain of retrovirus that is resistant to 2',3'-dideoxynucleoside analogues in appropriate target cells; however, the emergence of drug-resistant mutants must always be considered possible. It is important to stress that the phosphorylation reactions crucial to the activation of the nucleoside analogues are catalysed by host cell kinases, and therefore caution must be used in extrapolating from cells of one type (or species) to another.

Fig. 5 Inhibition of the viral cytopathic effect by thymidine congeners. *a*, By 3'-azido-3'-deoxythymidine; *b*, by 2',3'-dideoxythymidine; *c*, by the normal nucleoside, thymidine. Experimental protocol as in Fig. 3: solid bars, virus-treated and open bars, control cells.



If these kinases are lacking in the host cell, the retrovirus will appear resistant to the nucleoside analogues, but if the retrovirus is permitted to replicate in a different target cell which has the appropriate enzymes for anabolic phosphorylation, it will appear sensitive again. Similarly, we have observed that one dideoxynucleoside (2',3'-dideoxythymidine) behaves as a relatively poor substrate inhibitor for human thymidine kinase ($K_i > 500 \mu\text{M}$). The substitution of an azido group at the 3'-carbon of this analogue, yielding AZT, produces a compound that is an excellent substrate for thymidine kinase ($K_m 3 \mu\text{M}$) and is a very potent inhibitor of HTLV-III replication (Fig. 5*a*), of which more below. In this context, the lack of activity against HTLV-III that was observed using 2',3',5'-trideoxyadenosine (Fig. 2*e*) probably relates to the fact that the 5'-site is unavailable for phosphorylation. To re-emphasize, the retrovirus does not provide its own enzymes for phosphorylation; therefore, unlike herpes viruses, HTLV-III cannot adopt a strategy of mutating a kinase gene to develop drug resistance.

There are data to suggest that the HTLV-III reverse transcriptase is much more susceptible to the inhibitory effects of dideoxynucleotides as triphosphates than is mammalian DNA polymerase α (H.M., Z. Spigelman, R. P. McCaffrey and S.B., unpublished data), an enzyme that has key DNA synthetic and repair functions in the life of a cell. Indeed, reverse transcriptase has a higher affinity for dideoxynucleotides than for normal substrates. This parallels what had been learned in animal retroviral systems (see discussion in refs 46 and 58). Although 2',3'-dideoxynucleoside-5'-triphosphates can inhibit mammalian DNA polymerase β (a repair enzyme), and DNA polymerase γ (a mitochondrial enzyme)⁵⁸, we have observed that dideoxynucleosides can suppress HTLV-III replication and protect sensitive helper/inducer T-cell target cells *in vitro* for long periods without interfering with the function and survival of target T cells⁵⁵. As discussed above, our working explanation for the activity of these drugs against pathogenic retroviruses is that following anabolism to nucleoside-5'-triphosphates, they bring about a selective chain termination as the RNA form of the virus attempts to make DNA copies of itself. The inherent RNase H activity of the viral DNA polymerase or cellular RNases would be expected to degrade the viral RNA, giving the virus no second chance. Thus, one explanation for the selective antiretroviral activity of these compounds is that the viral enzyme is more easily fooled into accepting a dideoxy-

nucleotide than is its mammalian counterpart. Alternatively (or additionally) the virus may have less capacity than the host cell to repair the incorporation of the false nucleotide. In some cases, however, such drugs may perturb normal nucleoside metabolism or exert other effects that result in drug-related cellular toxicity. We will return to this point in the specific context of AZT, a special congener of dideoxythymidine.

AZT

Although synthesized over twenty years ago by Horwitz *et al.*⁴⁸ and shown to inhibit C-type murine retrovirus replication *in vitro* by Ostertag *et al.* more than 12 years ago⁵⁹, no medical application of AZT emerged until recently. As shown in Fig. 5*a*, AZT is a very potent *in vitro* inhibitor of the replication of HTLV-III and its cytopathic effect in susceptible target T cells, and it has an antiretroviral effect against widely divergent strains of HTLV-III¹³. The drug undergoes anabolic phosphorylation in human T cells to a nucleoside-5'-triphosphate⁶⁰, which can compete with thymidine-5'-triphosphate (TTP) and serve as a chain-terminating inhibitor of HTLV-III reverse transcriptase.

In that sense, AZT parallels the other dideoxynucleosides. But it can also bring about several alterations of nucleoside metabolism within host cells, which may be of clinical significance. For example, thymidine kinase catalyses the formation of a monophosphate derivative of AZT, which then serves as a competitive inhibitor, with low V_{max} , of thymidylate kinase (an enzyme that catalyses the second phosphorylation step to thymidine diphosphate), leading to the reduced formation of the third phosphorylation product, thymidine triphosphate (TTP)⁶⁰. From this perspective, AZT has an interesting capacity to lower the level of its natural substrate competitor (TTP) for viral and cellular DNA polymerases. AZT can also reduce 2'-deoxycytidine triphosphate (dCTP) levels⁶⁰, but the mechanism of this effect is unknown. How these reductions of normal pyrimidine pools might affect the balance between antiretroviral activity and drug-induced cytotoxicity also remains unknown. Nevertheless, pyrimidine starvation is likely to contribute to bone marrow suppression, one of the key side effects of the drug. This might be overcome by regimens that combine AZT with an agent that does not appear to deplete intracellular pyrimidine pools, such as dideoxycytidine. Other features which might have substantial clinical relevance, but remain unexplained at this time, include the capacity of aciclovir (a

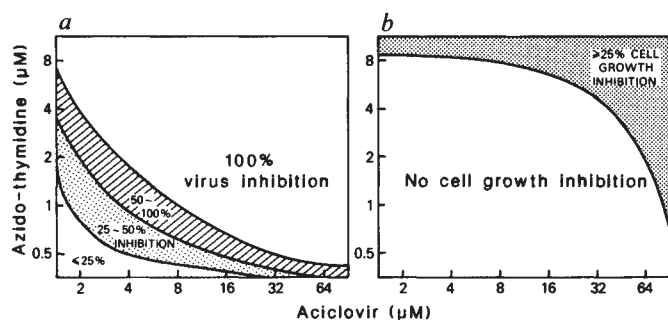


Fig. 6 Synergistic antiretroviral effect of AZT and aciclovir. ATH8 cells (2×10^5) were exposed to a very potent preparation of HTLV-III and cultured in the presence of various amounts of AZT and aciclovir. On day 6, viable cells were counted and the percentage of virus inhibition (a) (protective effect of the drugs on the survival of ATH8 cells exposed to the virus) and cell growth inhibition (b) were determined¹⁵. Note that although aciclovir alone showed rather weak antiretroviral effect at $\leq 64 \mu\text{M}$, combinations of both drugs exhibited a synergistic antiretroviral effect at a wide range of concentrations without affecting the growth of cells. In a, cross-hatched area represents 50-100% virus inhibition; dotted area 25-50% inhibition. Dotted area in b denotes $\geq 25\%$ cell growth inhibition.

guanosine analogue with potent activity against herpes viruses but rather weak antiretroviral activity) to potentiate the antiretroviral effect of AZT *in vitro* (Fig. 6).

We began administering AZT to patients with AIDS and related conditions in a phase I (feasibility and toxicity-testing) trial in July 1985. That study showed that the drug had good oral bioavailability and excellent penetration across the blood-brain barrier. Our conclusion from the phase I study¹⁴ and extensions of that study (R. Yarchoan and S.B., unpublished data) is that during short courses of administration AZT can bring about partial restoration of immune function and improve certain other clinical abnormalities associated with AIDS. The limiting toxicity observed was bone marrow suppression, predominantly in the form of anaemia and neutropenia. In some patients with HTLV-III associated dementia, AZT brought about quite prompt and dramatic neurologic improvement; however, the durability of the improvement is still not known⁶¹.

AZT was recently tested in a randomized study of patients with fulminant AIDS and poor prognostic AIDS-related complex under double-blind, placebo-controlled conditions at several centres. The study, which was begun in February 1986, had about 140 patients in each group. By September 1986 there were 19 deaths in the placebo group and only one in the AZT group ($P < 0.001$ by Cox's regression model). AZT-treated patients showed other evidence of clinical and immunological improvements. At that point, an independent data-monitoring board recommended that the placebo group begin to receive the drug. The study is still under way (with both groups now receiving drug) and the long-term safety and efficacy are now being assessed. Nevertheless, the results at hand warrant the initiation of appropriate controlled studies to see if therapy with AZT can block the onset of AIDS in individuals who have recently been infected by HTLV-III but who are otherwise free of symptoms.

Thus, AZT, a drug chosen on the basis of its selective *in vitro* antiviral effect against HTLV-III, has been shown to confer a clinical benefit in patients with advanced disease. In this respect, it represents a first step in developing practical chemotherapy against pathogenic human retroviruses. Although the clinical results might be important in their own right, they are also important because they appear to validate the idea that therapy for established retroviral infections is a feasible goal.

Conclusion

The genome of the virus that causes AIDS contains the standard replicative genes found in other retroviruses; however, it contains several additional genes. The product(s) of each gene represents a potential target in developing experimental therapies for diseases caused by this virus, and there are several steps related to the viral lifecycle which may be relevant for future therapeutic strategies. Curative therapy for diseases caused by pathogenic retroviruses will probably not be possible until the molecular biology of the virus and the structural chemistry of the key viral products are defined through further basic research, but it is not necessary to await new data before implementing some strategies that we already know have clinical applications in patients with advanced disease.

We thank Marvin Reitz, Flossie Wong-Staal and Dani Bolognesi for their helpful comments, and David Barry for his helpful discussions in our studies of the combinations of aciclovir and AZT, and for generously supplying these drugs.

- Gottlieb, M. S. *et al.* *New Engl. J. Med.* **305**, 1425-1431 (1981).
- Broder, S. & Gallo, R. C. *New Engl. J. Med.* **311**, 1292-1297 (1984).
- Gallo, R. C. *et al.* *Science* **224**, 500-503 (1984).
- Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497-500 (1984).
- Barré-Sinoussi, F. *et al.* *Science* **220**, 868-871 (1983).
- Levi, J. A. *et al.* *Science* **225**, 840-842 (1984).
- Coffin, J. *et al.* *Nature* **321**, 10 (1986).
- Shaw, G. M. *et al.* *Nature* **321**, 177-182 (1985).
- Ho, D. D. *et al.* *New Engl. J. Med.* **313**, 1493-1497 (1985).
- Resnick, L. *et al.* *New Engl. J. Med.* **313**, 1498-1504 (1985).
- Gartner, S. *et al.* *Science* **233**, 215-219 (1986).
- Koenig, S. *et al.* *Science* **233**, 1089-1093 (1986).
- Mitsuya, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 7096-7100 (1985).
- Yarchoan, R. *et al.* *Lancet* **i**, 575-580 (1986).
- Mitsuya, H., Matsukura, M. & Broder, S. in *AIDS: Modern Concepts and Therapeutic Challenges* (ed. Broder, S.) 303-333 (Dekker, New York, 1987).
- Gross, L. *Oncogenic Viruses* 3rd edn (Pergamon, Oxford, 1983).
- Baltimore, D. *Nature* **226**, 1209-1211 (1970).
- Temin, H. & Mizutani, S. *Nature* **226**, 1211-1213 (1970).
- Toh, H., Hayashida, H. & Miyata, T. *Nature* **305**, 827-829 (1983).
- Ratner, L. *et al.* *Nature* **313**, 277-284 (1985).
- Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* **40**, 9-17 (1985).
- Sodroski, J. *et al.* *Science* **227**, 171-173 (1985).
- Sodroski, J. *et al.* *Nature* **319**, 555-559 (1986).
- Sodroski, J. *et al.* *Nature* **321**, 412-417 (1986).
- Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. *Cell* **46**, 807-817 (1986).
- Dalgleish, A. G. *et al.* *Nature* **312**, 763-767 (1984).
- Klatzmann, D. *et al.* *Nature* **312**, 767-768 (1984).
- Pert, C. B. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 9254-9258 (1986).
- Sodroski, J., Goh, W. C., Rosen, C., Campbell, K. & Haseltine, W. A. *Nature* **322**, 470-474 (1986).
- Lifson, J. D. *et al.* *Nature* **323**, 725-728 (1986).
- Fisher, A. G. *et al.* *Science* **233**, 655-659 (1986).
- Matsushita, S. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 2672-2676 (1986).

- Veronese, F. D. *et al.* *Science* **231**, 1289-1291 (1986).
- Tanese, N., Sodroski, J., Haseltine, W. A. & Goff, S. P. *J. Virol.* **59**, 743-745 (1986).
- Mitsuya, H. *et al.* *Science* **226**, 172-174 (1984).
- Broder, S. *et al.* *Lancet* **ii**, 627-630 (1985).
- Rosenbaum, W. *et al.* *Lancet* **i**, 450-451 (1985).
- Sandstrom, E. G., Kaplan, J. C., Byington, R. E. & Hirsch, M. S. *Lancet* **i**, 1480-1481 (1985).
- Cullen, B. R. *Cell* **46**, 973-982 (1986).
- Okamoto, T. & Wong-Staal, R. *Cell* **47**, 29-35 (1986).
- McCormick, J. B., Getchell, J. P., Mitchell, S. W. & Hicks, D. R. *Lancet* **ii**, 1367-1369 (1984).
- Zamecnik, P. C., Goodchild, J., Taguchi, Y. & Sarin, P. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4143-4146 (1986).
- Henderson, L. E., Krutzsch, H. C. & Oroszlan, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 339-343 (1983).
- Schultz, A. M. & Oroszlan, S. *J. Virol.* **46**, 355-361 (1983).
- Ho, D. D. *et al.* *Lancet* **i**, 602-604 (1985).
- Mitsuya, H. & Broder, S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1911-1915 (1986).
- Robins, M. J. & Robins, R. K. *J. Am. chem. Soc.* **86**, 3585-3586 (1964).
- Horwitz, J. P., Chua, J. & Noel, M. *J. org. Chem.* **29**, 2076-2078 (1964).
- Doering, A. M., Janse, M. & Cohen, S. S. *J. bact.* **92**, 565-574 (1966).
- Horwitz, J. P., Chua, J., Noel, M. & Donatti, J. T. *J. org. Chem.* **32**, 817-818 (1967).
- Atkinson, M. R., Deutscher, M. P., Kornberg, A., Ruseil, A. F. & Moffat, J. G. *Biochemistry* **8**, 4897-4904 (1969).
- Toji, L. & Cohen, S. S. *J. Bact.* **103**, 323-328 (1970).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Cooney, D. A. *et al.* *Biochem. Pharmac.* **35**, 2065-2068 (1986).
- Mitsuya, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Chi, I.-M. *et al.* *Nature* **317**, 366-368 (1985).
- Dahiberg, J. E., Mitsuya, H., Broder, S. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Wagar, M. A., Evans, M. J., Manly, K. F., Hughes, R. G. & Huberman, J. A. *J. cell Physiol.* **121**, 402-408 (1984).
- Ostertag, W. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **71**, 4980-4985 (1974).
- Furman, P. A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 8333-8337 (1986).
- Yarchoan, R. *et al.* *Lancet* **i**, 132-135 (1987).

Mass distribution in spiral galaxies

L. K. Erickson, S. T. Gottesman & J. H. Hunter Jr

Department of Astronomy, University of Florida, Gainesville, Florida 32611, USA

The masses of a carefully selected sample of primary disk galaxies orbited by one or more satellite systems have been calculated both from the rotation of the disks and from the kinematics of the satellites. The ratio of these two mass estimates (χ_{obs}) is sensitive to the presence of any extended halo and for satellite-primary pairs the mass distribution is significantly different from that of binary galaxies. Our data cannot be described by the heavy halo model with circular orbits, though extended haloes are a possibility if the satellite orbits are highly eccentric.

PERHAPS the most important questions of contemporary, extragalactic astronomy concern the dark haloes that may surround disk galaxies. A popular view is that a typical halo is considerably more massive than the underlying disk, and that it extends out to several Holmberg radii¹. However, we believe that the observational basis for this surmise is not compelling. Massive haloes are not a necessary ingredient to produce flat rotation curves in disks. Van der Kruit² has discussed alternative methods and Hunter *et al.*³ have considered in detail the effect of truncating a disk. Also, massive haloes are not required to maintain stable disks⁴. While it seems likely that much unseen matter must reside in rich clusters of galaxies⁵, we believe the conclusion for small groups and for binary galaxies is less certain.

For small groups and binary galaxies, the observations provide radial velocities and angular separations of suspected members, which are assumed to be gravitationally bound. However, since the projection factors and orbital elements are unknown, the observations must be interpreted in a statistical fashion. Moreover, in the case of binary galaxies, the mass ratio of the members (and, hence, the location of the barycentre) usually is uncertain. Also, it is not clear that the members can always be treated as point masses. Recently, van Moorsel⁶ observed members of a carefully selected sample of binary, disk galaxies in H I, with a precision of $\approx 10 \text{ km s}^{-1}$. Since his observations provided a 21-cm rotation curve for each galaxy, the individual interior rotation masses could be calculated. Then, assuming that the orbital inclinations were distributed randomly, van Moorsel projected a large number of model pairs upon the plane of the sky, incorporating selection effects and a variety of assumptions about the orbital eccentricities, separations and the amount of dark matter associated with each disk. He concluded that his observations were best fit by models in which the mass of each halo is ~ 5 times that of its underlying disk, a result which he found to be insensitive to the orbital eccentricities and separations. We have carried out a similar analysis of the Magellanic Cloud-like satellites surrounding several large, disk galaxies and find less certain results.

Method

In observations⁷ of H I emission from the barred spiral galaxy NGC3992 the disk of the galaxy was well resolved at a high signal-to-noise ratio, and in addition, emission was detected from three small satellite systems. This makes it possible to determine the mass of NGC3992 by analysis of the kinematics of the disk (the rotation mass) as well as from an analysis of the motion of the satellites (the orbital mass). In our observations, the orbital mass measures the distribution of matter in a volume whose radius is ~ 3.5 times that of the disk. Therefore, if a galaxy has a significant and extended halo, there should be a pronounced difference between the orbital and the rotation masses. In essence, the satellites are assumed to be test particles

orbiting a central mass. In principle this method is superior to conventional studies of binary galaxies, because our satellites typically have $\sim 2.5\%$ of the mass of the primary and in all cases $< 8\%$. (This has been determined from the H I, global profiles of the satellites.) Hence, interaction effects should be negligible. We are examining the gravitational potential field of an undisturbed primary system, and in some cases we are exploring this potential simultaneously with several independent probes. The method is superior to optical mass estimation techniques, owing to the precision of the radio velocity measurements ($\epsilon \approx 10 \text{ km s}^{-1}$). Unlike the studies of Peterson⁸, or of Helou *et al.*⁹, the rotation law is observed directly. Following the suggestion of Lequeux¹⁰, the 'rotation mass' can be estimated from a keplerian model ($M = rv^2/G$). By definition, this estimate includes the gravitational effects of all the mass (disk and halo) within the radius of the most distant observed point in the H I disk. This procedure may overestimate M , for were the mass confined to a thin disk, the total mass would be in the range 0.6–0.76 of the spherical distribution^{6,10}. The keplerian mass was used in our calculations.

In our analysis, we have adopted a method described by van Moorsel⁶, in which each group is treated as a multiple binary system. Then, we may write the orbital mass as,

$$M \cdot \chi(\nu, i, \omega, e) = \frac{rV_r^2}{G} \quad (1)$$

with

$$\chi = \frac{\sin^2 i}{1 + e \cos \nu} [\cos(\nu + \omega) + e \cos \omega]^2 \times [1 - \sin^2(\nu + \omega) \sin^2 i]^{1/2} \quad (2)$$

where r and V_r are the projected radial separation and radial velocity difference. The function χ depends upon the true anomaly, ν , the angle between the line of nodes and the major axis of the orbit, ω , the inclination of the orbital plane, i , and e , the eccentricity of the orbit. As a measure of the amount of dark matter we have

$$\chi_{\text{obs}} = \frac{rV_r^2}{G(m_1 + m_2)} \quad (3)$$

where m_1 and m_2 are the disk masses of the binary galaxies. In our case $m_2/m_1 \ll 1$ and, hence, m_2 can be ignored and the subscripts dropped. χ_{obs} is related to χ by

$$\chi_{\text{obs}} = \frac{M}{m} \chi \quad (4)$$

From a Monte-Carlo analysis, we have verified van Moorsel's conclusion that the probability distribution, $P(\chi)$, is peaked sharply for small values of χ . In other words, there is a strong tendency to underestimate the orbital mass (by almost an order of magnitude), because small projected radii and velocities are

Table 1 Primary groups

System	Galaxy type*	V_{sys}^+ (km s ⁻¹)	Distance to primary galaxy (Mpc)	Inclination angle of primary disk (deg.)	R_{max} (kpc)	$V(R_{\text{max}})$ (km s ⁻¹)	$m(V)^\ddagger$ ($\times 10^{10} M_\odot$)	mag	Notes
NGC224	SAS3 SbI-II	-201	0.7	78	28.0	230	34.4	4.3	Rotation curve from CO measurements ²² , optical velocities from ref. 23.
NGC1023	SBO- SBO1(5)	610	7.5	80	9.5	251	13.9	10.5	Rotation curve from ref. 24, H I measurements from ref. 13.
NGC1961	SXT5 Sb(rs)IIpec	3935	41.3	50	24.0	400	75.0	12.2	H I measurements and rotation curve from ref. 14; mass calculations from ref. 12.
NGC3359	SBT5 SBc(s)I.8pec	1009	11.0	51	20.8	140	9.5	11.0	H I measurements from ref. 25.
NGC3992	SBT4 SBb(rs)I	1046	14.2	53	19.0	240	25.4	10.7	H I measurements and calculations from ref. 18.
NGC3992	SBT4 SBb(rs)I	1046	14.2	53	19.0	240	25.4	10.7	H I measurements and calculations from ref. 18.
NGC4303	SXT4 Sc(s)I.2	1556	12.9	29	12.2	160	7.3	10.9	H I measurements from refs 27 and L.K.E., distance of group from ref. 28.
NGC4731	SBS6 SBc(s)III	1490	10.5	54	13.0	150	6.8	6.0	H I measurements and calculations from ref. 18
NGC5084	L(-2) SO1(8)	1721	15.0	>86	34.0	328	85.0	—	H I measurements and calculations from ref. 17.

* Morphological type of the primary galaxy, from refs 20, 21.

† The measured heliocentric systemic velocity of the primary galaxy.

‡ Keplerian disk mass calculation using $m = V(R_{\text{max}})^2 R_{\text{max}} / G$.

§ The apparent blue magnitude from the Uppsala catalogue¹¹.

highly favoured. On the assumption that the rotation law of the primary reflects all of its mass, and that the system is bound, $P(\chi)$ should fall to zero when $\chi = (1 + e) \leq 2$.

Selection criteria

Thus, our approach to the general problem of the mass distribution in spiral galaxies is to observe large spiral galaxies (primaries) surrounded by one or more relatively massless satellites. These small groups are found by systematically searching the Uppsala General Catalog¹¹. The primaries must be large enough to be resolved at λ 21 cm by the Very Large Array (VLA) telescope and the primaries and satellites have to be rich enough in atomic hydrogen that their kinematic properties can be established with high precision. Furthermore, we desire (primary + satellite) systems which are compact enough in their distribution that they can be observed in one field (diameter ≈ 30 arc min) by the VLA. Finally, by restricting the velocity differences within the groups, we hope to have eliminated most of the problems associated with contamination by field galaxies. The dominance of the primary galaxy is ensured by requiring that it be two magnitudes brighter than, or 1–2 arc min larger than, any other member of the group. We have rejected satellites with velocity of differences ≥ 600 km s⁻¹, a limit imposed by our receivers. (Van Moorsel found that such systems mostly were non-physical pairs.) Six groups surrounding the primary galaxies were indicated by our selection criteria: NGC1961, NGC3992, NGC4111, NGC4258, NGC4303 and NGC3893. The first two were considered earlier, for other reasons, by Gottesman *et al.*¹² and by Gottesman and Hunter⁷.

Results

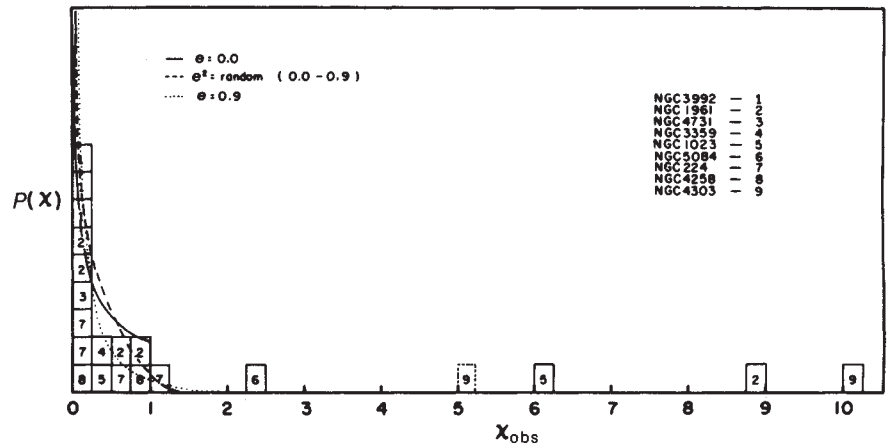
We have augmented our results from the small groups by adding galaxies that we have observed previously, which possess only single satellites (NGC3359, NGC4731 and NGC5084). Also, we have taken information from the literature on M31 and NGC1023. The data for the primaries and their disk masses are given in Table 1, and the data for the satellites and the calculations of χ_{obs} are shown in Table 2. In Fig. 1 we have plotted the distribution of χ_{obs} . The diagram displays 20 primary-satel-

lite pairs (which is four more than the number of binary systems of van Moorsel); of these only four have $\chi_{\text{obs}} > 2.0$. By comparison, more than half of van Moorsel's binaries showed $\chi_{\text{obs}} > 2.0$. Therefore, he concluded that his disk systems contain haloes with masses typically five times their rotation masses. This is not the conclusion we would draw from Fig. 1. With the exception of the four groups we discuss below, the χ_{obs} distribution can be described by a population of bound satellites. We show in Fig. 1 model distributions for various orbital eccentricities. The theoretical and observed distribution were compared using a Kolmogorov-Smirnov test. The distributions are the same at better than the 5% confidence level and, therefore, no masses other than the rotation masses of the primaries are needed to bind these systems. As we are considering relatively massless satellites, we believe our technique is more risk-free than that of van Moorsel. Both van Moorsel and we are examining the consequences of point source, two-body dynamics, a situation more closely realized in our context than in his. Nonetheless, without further examination, the explanation for the apparent disagreement between the two results is not obvious.

NGC1023 group. Sancisi *et al.*¹³ believe that this is an SO that has recently undergone a tidal encounter. The mass of the system from the 'tidal rings' is $6.5 \times 10^{11} M_\odot$ ($R = 45$ kpc) compared to $1.8 \times 10^{11} M_\odot$ within the optical dimensions ($R = 12.6$ kpc). The two dwarfs listed in Table 2 are well isolated from any tidal phenomena, and, as no inclination corrections have been applied to the disk masses, our χ_{obs} values are upper limits. Thus, bearing in mind the unusual nature of this galaxy, it appears to have a more extended mass distribution than the optical results would indicate.

NGC1961 group. This group was studied by Shostak *et al.*¹⁴ using the Westerbork synthesis telescope. Gottesman *et al.*¹² used the satellites to measure the mass of the primary. The satellite UGC3349 indicates a χ_{obs} of ~ 6 , where we have used the smaller of the two mass estimates from Shostak *et al.*¹⁴. This may indicate a significant halo for the primary. However, we caution that NGC1961 is a peculiar system. The group appears to be embedded in an intergalactic medium which, it is suggested, is stripping the H I from NGC1961. The presence of this

Fig. 1 The theoretical distribution of χ compared with the observed distribution (χ_{obs}). Each rectangle represents one primary-satellite pair for which the primary is identified. The dotted value for NGC4303-UGC7439 is taken from van Gorkom's data. Three model distributions are shown with different orbital eccentricities. The distributions are normalized to the same area.



extended, low density component may complicate the dynamical analysis.

NGC4303 group. This group is made up of three satellites in the field of the VLA telescope. Our observations, which are in progress, show a strong detection for UGC7439 and a possible detection of UGC7404 which we have not listed in Table 2, as it is in conflict with a measured optical value. Regardless, the satellite velocities are large and, consequently, the orbital mass is large too. Our observed maximum velocity in the disk is very small, $\sim 79 \text{ km s}^{-1}$ at a radius of $\sim 11 \text{ kpc}$. Thus, a substantial correction for the inclination may have to be made. J. van Gorkom (personal communication) has allowed us to use her data, with better angular and velocity resolution than ours. We find that the inclination could be as small as 17° , less than

our data indicate. This inclination is within the range of values quoted by Huchtmeier *et al.*¹⁵ of $14^\circ < i < 48^\circ$. Also, Rubin *et al.*¹⁶ predict a maximum rotation velocity of 209 km s^{-1} for a typical Sc of luminosity class I.1. This would imply an inclination of 18° from van Gorkom's data. Table 2 lists the values from our data, but even if the maximum rotation velocity is 225 km s^{-1} (van Gorkom's data) $\chi_{\text{obs}} = 5.1$ for UGC7439. This would imply the presence of a substantial dark halo. However, if the inclination is closer to 48° rather than 14° , the secondaries may not be bound. Also it is possible that the environment of the Virgo cluster adds to the dynamical complexities of this small group.

NGC5084 group. The observations were made by Gottesman and Hawarden¹⁷. The primary is a peculiar, massive SO^+ object with an anomalous gas/dust annulus. The calculated χ_{obs} value

Table 2 Satellites

Group satellite	Heliocentric velocity (km s^{-1})	Measurement method	del V (km s^{-1})*	Separation of satellite and galaxy (kpc)	$M(\text{del V})$ ($\times 10^{10} M_\odot$)†	χ_{obs} (M/m)	mag‡
NGC224	-301	H I					4.3
NGC147	-168	opt	142	90.2	42.3	1.23	12.0
NGC185	-208	opt	100	85.6	20.0	0.58	11.0
NGC205	-240	opt	62	7.5	0.7	0.02	9.4
NGC221	-216	opt	84	4.9	0.8	0.02	9.2
NGC1023	610	H I					10.5
North	905	H I	295	34.2	85.9	6.18	—
South	695	H I	85	17.5	3.7	0.26	—
NGC1961	3,935	H I					12.2
A	4,108	H I	173	91	62.6	0.84	—
B	3,895	H I	40	106	3.9	0.05	—
B1	3,800	H I	135	122	52.4	0.70	—
UGC3342	3,927	H I	8	157	0.2	0.00	15.4
UGC3349	4,282	H I	347	240	671.9	8.96	14.4
NGC3359	1,009						11.0
dwarf	962	H I	47	48.5	2.5	0.26	—
NGC3992	1,046	H I					10.7
UGC6923	1,062	H I	16	60.6	0.4	0.02	14.1
UGC6940	1,112	H I	66	35.4	3.6	0.14	16.0
UGC6969	1,115	H I	69	45.3	5.0	0.20	15.5
NGC4258	465						9.6
UGC7356	255	opt	210	18.8	19.3	0.99	17.0
UGC7335	490	H I	25	19.1	0.3	0.02	13.9
NGC 4303	1,556						10.9
UGC7439	1,270	H I	286	39.1	74.4	10.19	14.9
NGC4731	1,490	H I					6.0
RNGC4731a	1,505	H I	15	32.1	0.2	0.03	—
NGC5084	1,721	H I					—
dwarf	2,089	H I	368	65.6	206.5	2.43	—

* The absolute value of the velocity difference between the primary systemic velocity and the satellite velocity; this is the quantity V_r in equation (1). Note that the NGC224 velocities are in the ASR reference frame, not heliocentric.

† The orbital mass of the primary galaxy computed from $M = (\text{del V})^2 R / G$, with R the separation between the primary and satellite galaxies.

‡ The apparent blue magnitudes of the members, from the Uppsala catalogue¹¹.

of ~ 2.4 is at the limit of significance considering the uncertainty in the extreme inclination of the disk, and of the systemic velocity of the satellite. (The global H I velocity profile of the satellite is not symmetrical.)

NGC3992 group. This group has been discussed extensively by Gottesman and Hunter⁷ and by Gottesman *et al.*¹⁸. The disk mass we have used comes from a keplerian calculation at $R_{\max}$. Thus, the mass differs slightly from that given by Gottesman *et al.* None of the observations reveal the presence of a massive halo. Dynamical modelling of NGC3992 by Hunter *et al.*¹⁹ leads to similar conclusions. We believe that the three independent analysis of the data indicate most strongly that this system does not have a significant halo. Indeed, it may be the best example of such a spiral galaxy.

Thus, from our observations, at most four groups have χ_{obs} values >2 . Hence, the primary galaxies in these systems may contain massive haloes. In contrast NGC3992 almost certainly does not have a massive halo. What requires further interpretation is the observed distribution of χ .

Discussion

Van Moorsel found that he could bind his pairs if he assumed a model in which the flat rotation curves of both galaxies were extended until the mass distributions touched. We have considered models of this type. We assume the satellites to be massless points moving in the potential of the halo of the primary. The mass of the primary is obtained by extrapolating a flat rotation curve out to the distance of each satellite. ($M \propto R$; halo density $\propto R^{-2}$.)

Following van Moorsel, we write,

$$\chi' = \frac{R}{r} \chi = \left[\frac{V_r}{V(R_{\max})} \right]^2 \quad (5)$$

where R is the true separation of the satellite-primary pair, and r is the observed separation. Assuming the data are unbiased, the hypothesis that the model correctly describes the observations can be rejected. That is, we conclude from a Kolmogorov-Smirnov test that van Moorsel's simple halo model, involving circular orbits, will not reproduce our observations. There are too many small values of χ . However, this model may not be appropriate if violent collapse of the haloes of the primaries caused the satellites to have radial orbits. We have not explored this situation exactly, for if the mass of the primary is proportional to r then the gravitational force on a satellite is proportional to r^{-1} , and the orbits must be recalculated. As we are sampling the satellites at different locations in their orbits we have approximated the real situation by a simple scaling of the (point) orbital mass appearing in χ . The scaling factor is an adopted value of R/r . This ratio ranges from 0.2 to 12, for our data, with an average value of 3.5.

The shortcoming of this model is that it greatly overcalculates the forces and velocities near pericentre but not at apocentre. Thus, χ values predicted by the model are too large at small separations. However, more detailed modelling is warranted because we find that χ does not correlate with projected separation. We conclude that with the exception of the four unusual systems we have discussed, the χ_{obs} distribution can be modelled successfully by galaxies whose masses equal the keplerian mass defined by the last observed point in the disk. No massive haloes are required to bind the satellites. In addition, our data cannot fit van Moorsel's halo model with circular orbits, and as R/r increases, satisfactory models only can be constructed by assuming increasingly eccentric orbits for the satellites (for $R/r = 2.5$, $e \geq 0.6$; for $R/r = 3.5$, $e \geq 0.9$; for $R/r = 5$, $e > 1$).

While this approach is not completely satisfactory, it does represent an upper limit, for the satellites feel the full effect of the mass of the primary at all locations in their orbits. Thus, our data probably cannot accommodate halo masses in excess of ~ 3 times their keplerian masses, and such an increase in the masses of the primaries requires the satellite orbits to be very eccentric.

If the model presented here is correct, then NGC3992 is unusual in that it does not have an extended halo. In contrast, NGC1023, NGC1961, NGC4303, and perhaps NGC5084, either must have massive haloes or be unbound systems.

Throughout this discussion we have presumed that our data sample is unbiased. However, at least three of our primaries are strongly barred and our χ_{obs} distribution may not reflect a homogeneous sample of disk galaxies. Also, the χ_{obs} distribution may reflect selection effects in our database. Other differences between van Moorsel's sample and ours are that he explored a larger volume (relative to a typical disk) than we have, and our groups cover a wider range of separations. Van Moorsel found χ_{obs} to be very insensitive to the applied selection criteria. However, as we chose galaxy pairs with larger magnitude differences, our data may be biased differently. Following van Moorsel's precepts, we have created a synthetic UGC¹¹ and from this we have modelled our selection effects. Like van Moorsel we find a pronounced difference between the biased and unbiased distribution of galaxy separations. However, again like van Moorsel, we find, that there is no significant difference in the mean values of χ for the two distributions. We can only echo his conclusion; 'the distribution of the parameter χ is very insensitive to the applied selection'. Hence, the observed distributions are different.

We thank Carl Cole, Heinz Eichhorn and Haywood Smith for discussions. J. H. van Gorkom very kindly allowed us to make use of her data on NGC4303 before publication. Our observations were made at the National Radio Astronomy Observatory which is operated by AUI under contract to the NSF.

Received 1 August 1986; accepted 2 January 1987.

1. Faber, S. M. & Gallagher, J. S. *Astr. Astrophys.* **17**, 135-187 (1979).
2. van der Kruit, *Proc. astr. Soc. Aust.* **5**, 136-143 (1983).
3. Hunter, J. H., Ball, R. & Gottesman, S. T. *Mon. Not. R. astr. Soc.* **208**, 1-14 (1984).
4. Sellwood, J. A. in *Internal Kinematics and Dynamics of Galaxies* (ed. Athanassoula, E.) 197-202 (Reidel, Dordrecht, 1983).
5. Geller, M. J. in *Clusters and Groups of Galaxies* (eds Mardirossian, F., Giuricin, G. & Mezzetti, M.) 353-365 (Reidel, Dordrecht, 1984).
6. van Moorsel, G. A. thesis, Gröningen Univ. (1982).
7. Gottesman, S. T. & Hunter, J. H. *Jr Astrophys. J.* **260**, 65-69 (1982).
8. Peterson, S. *Astrophys. J.* **232**, 20-33; *Astrophys. Suppl. Ser.* **40**, 527-575 (1979).
9. Helou, G., Salpeter, E. E. & Terzian, Y. *Astr. J.* **87**, 1443-1464 (1982).
10. Lequeux, J. *Astr. Astrophys.* **125**, 394-395 (1983).
11. Nilson, P. *Uppsala General Catalog of Galaxies* (Uppsala Astron. Obs. Ann. vol. 6, 1973).
12. Gottesman, S. T., Hunter, J. H. Jr & Shostak, G. S. *Mon. Not. R. astr. Soc.* **202**, 21-24 (1983).
13. Sancisi, R., van Woerden, H., Davies, R. D. & Hart, L. *Mon. Not. R. astr. Soc.* **210**, 497-514 (1984).
14. Shostak, G. S., Hummel, E., Shaver, P. A., van der Hulst, J. M. & van der Kruit, *Astr. Astrophys.* **115**, 293-307 (1982).
15. Huchtmeier, W. K., Richter, O. G., Bohnstengel, H. D. & Hauschildt, M. ESO Preprint no. 250 (1983).
16. Rubin, V. C., Ford, W. K. & Thonnard, N. *Astrophys. J.* **238**, 471-487 (1980).
17. Gottesman, S. T. & Hawarden, T. G. *Mon. Not. R. astr. Soc.* **219**, 759-775 (1986).
18. Gottesman, S. T., Ball, R., Hunter, J. H. & Huntley, J. M. *Astrophys. J.* **286**, 471-490 (1984).
19. Hunter, J. H., Ball, R., Huntley, J. M., England, M. & Gottesman, S. T. *Astrophys. J.* (submitted).
20. de Vaucouleurs, G., de Vaucouleurs, A. & Corwin, H. G. Jr *Second Reference Catalog of Bright Galaxies* (University of Texas, Austin, 1976).
21. Sandage, A. & Tammann, G. A. *A Revised Shapley Ames Catalog of Bright Galaxies* (Carnegie Institution, Washington DC, 1981).
22. Roberts, M. S., Whitehurst, R. N. & Cram, T. R. *IAU Symp.* No. 77, 169-173 (1978).
23. Einasto, J. & Lynden-Bell, D. *Mon. Not. R. astr. Soc.* **199**, 67-80 (1982).
24. Dressler, A. & Sandage, A. *Astrophys. J.* **265**, 664-680 (1983).
25. Ball, R. *Astrophys. J.* **307**, 453-471 (1986).
26. van Albada, G. D. thesis, Gröningen Univ. (1978).
27. McCutchen, W. H. & Davies, R. D. *Mon. Not. R. astr. Soc.* **150**, 337-357 (1970).
28. de Vaucouleurs, G. *Astrophys. J.* **227**, 729-755 (1979).

Bipolar affective disorders linked to DNA markers on chromosome 11

Janice A. Egeland*, Daniela S. Gerhard^{†||}, David L. Pauls[‡], James N. Sussex*, Kenneth K. Kidd[§], Cleona R. Allen*, Abram M. Hostetter* & David E. Housman[†]

* Department of Psychiatry, University of Miami School of Medicine (D-29), PO Box 016960, Miami, Florida 33101, USA

[†] Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[‡] Child Study Center and Department of Human Genetics and [§] Departments of Human Genetics and Psychiatry, Yale University School of Medicine, New Haven, Connecticut 06510, USA

An analysis of the segregation of restriction fragment length polymorphisms in an Old Order Amish pedigree has made it possible to localize a dominant gene conferring a strong predisposition to manic depressive disease to the tip of the short arm of chromosome 11.

ESTABLISHING the role of genetic factors in the aetiology of mental illness has represented a formidable challenge. The separation of environmental from intrinsic biological factors and the complexities of psychiatric diagnosis are major obstacles in this endeavour. Nevertheless, evidence of biological and genetic contributions to aetiology make the major affective disorders excellent candidates to address this issue.

Major affective disorders are a group of illnesses manifested by disturbance in mood, and in physiological, cognitive and endocrine functions¹. The primary forms are unipolar major depression and bipolar disorder, distinguished by the occurrence in the latter of manic episodes, typified by an elevated, expansive or irritable mood. Symptoms of mania can include physical restlessness, increased talkativeness, racing thoughts, inflated self-esteem extending to delusions of grandeur, decreased need for sleep and distractibility. Manic episodes usually lead to impairment in both social and occupational function because of excessive behaviour patterns. Hypomania is less severe and has fewer symptoms. When mania or hypomania occurs in association with depressive episodes, the diagnosis is bipolar I (BP I) or bipolar II (BP II) disorder, respectively.

Depressive episodes are characterized by a mood of hopelessness. Associated symptoms are appetite and weight changes, sleep disturbance, difficulty in concentrating, decreased energy, loss of interest, and feelings of worthlessness or guilt, extending to thoughts of death or suicide. Major and minor depression are distinguished by the number and severity of these symptoms.

Bipolar disorder (manic-depressive illness) has an estimated lifetime prevalence of ~1% in North America and western Europe². Onset is usually between 15 and 35 years of age. Major depression has a much higher estimated prevalence with a wider range in age of onset². Both manic and depressive episodes can persist for periods of a week to several months and may require hospitalization.

The response of manic patients to lithium salts³ and of depressed patients to the tricyclic antidepressants, monoamine oxidase inhibitors and related medications^{4,5} indicates that neuronal systems affected by these agents are important in these illnesses. The biological basis of manic and depressive states also is evidenced by cycling of mood unrelated to specific situational factors and by altered patterns in energy, sleep and appetite. Twin studies, adoption studies and family studies demonstrate that genetic factors can play an important role in the aetiology of affective disorders⁶⁻¹⁰. However, establishing the mode of transmission and identification of a major genetic locus has been problematic. To evaluate the genetic contribution to major affective disorders, a study of the genetically isolated

Old Order Amish (OOA) population in southeastern Pennsylvania (the Amish Study¹¹) was initiated in 1976.

Amish study

The Amish study has presented a unique opportunity to examine hypotheses of genetic transmission. Detailed genealogical records document that all the 12,000 members of the OOA population are descended from 30 progenitors who emigrated from Europe in the early eighteenth century¹². Over the past 25 years this closed community has been studied extensively by one of us (J.A.E.). Large family size, clearly established paternity and geographic concentration made it excellent for epidemiological and pedigree studies. Social and religious prohibitions of alcohol and drug use permitted more accurate ascertainment and diagnosis of affective disorders. Prevalence of affective disorders among the OOA was found to be similar to the general North American population¹³. The characteristics of the major forms of affective disorder, including basic symptoms, age of onset, course of illness and response to pharmacological treatment, were indistinguishable for Amish and non-Amish patients.

To date, the study has focused on families of bipolar patients. Segregation analyses of these family data indicate that an autosomal dominant mode of transmission is most consistent with the patterns of illness observed in the kindreds. Results suggest that the allele predisposing individuals to bipolar affective disorders has a frequency of 0.021 ± 0.002 in the population¹². Penetrances were estimated using age-specific estimates of the population prevalence for the illness. The susceptible genotypes (AA and Aa) were estimated to have a maximum penetrance of 0.63 at age 30 and above. The 'normal' genotype (aa) was estimated to have penetrance 0.0001. These results provided a strong impetus to genetic linkage studies.

Initial linkage studies, including information from a number of polymorphic serum proteins and blood-group antigens did not give evidence for linkage to a major locus for affective disorders¹⁴. Recombinant DNA techniques provide a vehicle for dramatically expanding the number of markers available for genetic linkage studies through the identification of restriction fragment length polymorphisms (RFLPs)¹⁵. To implement this strategy in the Amish study, blood samples were collected from a multigenerational pedigree, highly informative for bipolar illness. Permanent lymphoid cell lines were established from these blood samples at the Coriell Institute for Medical Research¹⁶. DNA derived from 51 of these cell lines was used in an initial series of RFLP studies. Preliminary evidence was obtained suggesting possible linkage between a major locus for bipolar illness and two markers on chromosome 11, insulin (INS) and the cellular oncogene Ha-ras-1 (HRAS1)^{17,18}. To extend these studies, DNA samples were prepared from permanent cell lines established from 30 additional members of

|| Present address: Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA.

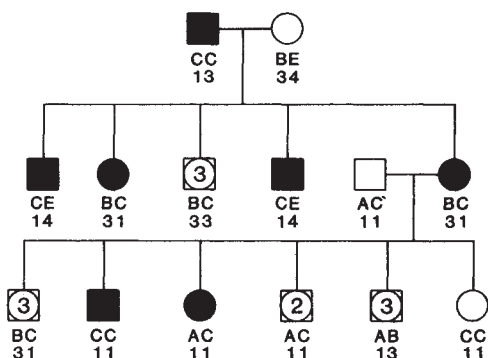


Fig. 1 *INS* and *HRAS1* typings on selected members of Pedigree 110. Both the *INS* and *HRAS1* loci have multiple alleles as a result of an insertion/deletion sequence in the immediate vicinity of the coding region^{24,25}. Four different alleles at the *INS* locus were observed when genomic DNA of the individuals shown in this figure was digested with the enzyme *PvuII*. The alleles are: A, ~760 base pairs (bp); B, ~800 bp; C, ~830 bp; D, ~2,300 bp. Alleles of *HRAS1* are as follows: 1, 5.2 kilobases (kb); 2, 5.7 kb; 3, 6.2 kb; 4, 6.8 kb; on *SacI*-digested DNA. Filled symbols, individuals with affective disorder; open symbols, individuals without affective disorder. Squares, males; circles, females; superimposed square and circle, other siblings (number given in centre). Note that alterations in sex designation and reductions in sibship size have been made to preserve anonymity of individuals at risk of affective disorder.

Methods. DNA from each individual was digested by either *SacI* (*HRAS1*) or *PvuII* (*INS*), electrophoresed on 0.7% Tris-acetate agarose gel (*HRAS1*) or 1.2% Tris-borate agarose gel (*INS*), transferred to Xetabind (nylon-based membrane, AMF-Cuno) by a method⁴⁴ modified from Southern. Filters were pretreated in 0.5% SDS, 0.1×SSC at 60 °C for 1 h and prehybridized overnight in 50% formamide, 6×SSC, 50 mM sodium phosphate pH 6.5, 5×Denhardt's solution, 100 µg ml⁻¹ sonicated, denatured salmon sperm DNA and 0.5% SDS. Hybridization was carried out in the above solution but with only 25 mM sodium phosphate and 1×Denhardt's solution for two days, using as probes the plasmid phins310 for *INS* and pEJ6.6 for *HRAS1*. Incubation temperature was 50 °C for *INS* and 42 °C for *HRAS1*. Filters were washed once at room temperature in 2×SSPE, 0.1% SDS, twice at 55 °C in 2×SSPE, 0.1% SDS and once at 55 °C in 0.1×SSPE, 0.1% SDS. SSC (1×) is 150 mM NaCl, 15 mM sodium citrate.

this pedigree¹⁶. We now report compelling evidence for tight linkage between two DNA sequences located on chromosome 11 and a locus conferring a strong predisposition to bipolar affective disorders.

Case ascertainment and diagnoses

Methods for case ascertainment included both a community-wide epidemiological survey among the Amish and a search of all Amish medical records at local psychiatric hospitals. Multiple interviews of patients and family members using the Schedule for Affective Disorders and Schizophrenia-Lifetime Version¹⁹ (SADS-L) and abstracts of hospital and clinic records formed the basis for diagnosis. Diagnostic evaluations were made by a psychiatric panel blind to patient identity, biological relationships, previous diagnoses and treatment response. When both interview and medical abstracts were available, they were evaluated separately. The panel adhered strictly to research diagnostic criteria²⁰ which define mood, symptoms, duration and severity. High diagnostic reliability was demonstrated²¹.

Thirty-two bipolar families were identified from a larger sample of psychiatrically ill patients. The linkage studies reported here are based on one large composite family (pedigree 110) which encompasses three of the 32 bipolar families originally ascertained. We focus on 81 members from pedigree 110 including 19 affected and 62 psychiatrically well individuals.

Of the 19 subjects with major affective disorders, 14 had some form of bipolar affective illness. Based on research diagnostic

criteria, 11 were diagnosed as bipolar I, one was diagnosed as bipolar II, one had a schizoaffective disorder, manic-depressive type and one had atypical psychosis with prominent affective features. Five members of pedigree 110 suffered from unipolar major depression. An additional five individuals were diagnosed with minor psychiatric conditions but were considered unaffected in the linkage analyses. Systematic screening methods were used to update changes in psychiatric status on a yearly basis for pedigree 110. Consequently, the diagnoses used in these analyses do not correspond completely to those originally published for this family¹⁶.

Genetic markers and linkage analysis

DNA samples from members of pedigree 110 were typed (blind to diagnosis) for segregation of a number of loci located on chromosome 11. The results for two loci close to the distal end of the short arm of the chromosome, *INS* and *HRAS1*, are presented here. Both these loci, located ~3 map units apart^{22,23} exhibit high frequency RFLPs due to the insertion or deletion of a short repeated DNA sequence in the non-coding region adjacent to the gene^{24,25}. The multiple alleles that are a consequence of these mutations were generally adequate to provide segregation information in most meioses. Because several individuals were homozygous for the insertion/deletion polymorphism, a *TaqI* site polymorphism outside the *INS* gene²⁶ was also typed. Haplotypes were then established for the insertion/deletion and the *TaqI* polymorphism. The resulting haplotypes were used in linkage analyses.

The segregation of the *INS* and *HRAS1* loci in a segment of pedigree 110 is shown in Fig. 1. In the two oldest generations, allele 1 of *HRAS1* and the C allele of *INS* cosegregate with the occurrence of bipolar affective disorder. All members of these two generations are past the upper end of the age-of-onset spectrum for bipolar disorders. In the third generation all affected individuals have also inherited the 1 allele at *HRAS1* and the C allele at *INS* from an affected parent. All well individuals who received this combination of alleles from an affected parent were still within the age of risk for developing the illness.

To determine the statistical significance of cosegregation of these markers with a proposed gene for predisposition to bipolar affective disorder, linkage analyses were performed using the computer program LIPED²⁷, modified to allow for age-dependent penetrance²⁸. The accuracy of these analyses is dependent upon correct assessment of clinical status as well as the specification of correct values for penetrances and gene frequen-

Table 1 Estimates of gene frequency and age-specific penetrance for affective disorder from the Amish study

Age	Autosomal dominant model			
	Total sample $q = 0.021 \pm 0.002$		Pedigree 110 $q = 0.015 \pm 0.005$	
	f_{A-}	f_{aa}	f_{A-}	f_{aa}
15-19	0.10	0.0000	0.18	0.0001
20-24	0.27	0.0001	0.42	0.001
25-29	0.56	0.0001	0.73	0.001
30+	0.63	0.0001	0.85	0.001

Quantities: q , frequency of the susceptibility allele A; f_{A-} , penetrance for genotypes AA or Aa; f_{aa} , penetrance for genotype aa. Segregation analyses were performed using the computer program POINTER⁴². The total sample included relatives of all 32 BP I probands ascertained from the OOA population. All relatives younger than 14 were excluded from the analyses. For these analyses, relatives with diagnoses of BP I, schizoaffective, BP II or major depressive disorder were included as affected. All other relatives were considered unaffected. Population prevalences incorporated were obtained from the entire OOA community. A wide range of genetic hypotheses were examined in an hierarchical fashion. Likelihood ratio tests were employed to determine statistical significance¹³.

Table 2 The lod scores for pairwise linkage analyses between *HRAS1* and a major affective disorders locus for four sets of penetrance values

Maximum penetrance	Recombination frequency (θ)						
	0.000	0.001	0.010	0.030	0.050	0.070	0.100
0.55	3.070	3.068	3.041	2.971	2.888	2.796	2.643
0.63	3.340	3.337	3.310	3.239	3.155	3.059	2.898
0.85	4.083	4.080	4.061	3.997	3.912	3.810	3.631
0.95	4.322	4.321	4.310	4.258	4.178	4.077	3.894

Pairwise linkage analyses were performed with the computer program LIPED²⁵ modified to include age-dependent penetrance. Analyses were performed assuming a dominant mode of transmission. The estimate of trait allele frequency was fixed at 0.021. Penetrance varied linearly from 0.0 at age 15 to the maximum at age 35 and beyond. Allele frequencies were estimated from a large reference sample as follows: allele 1, 0.3; allele 2, 0.3; allele 3, 0.1 and allele 4, 0.3. A wide range of penetrance estimates for genotypes *AA* and *Aa* were examined. The penetrance of *aa* in all analyses was fixed at 0.001. The values of penetrance used for *AA* and *Aa* included those obtained from segregation analysis of the total sample of 32 BP families and from pedigree 110 alone. In addition, penetrance estimates obtained from the inclusion of the standard error of the parameter estimates from segregation analyses were incorporated. The recombination fraction (θ) varied from 0.5 to 0.10 in increments of 0.05, and from 0.10 to 0 in the increments shown. A lod score represents the logarithm to the base 10 of the likelihood (*L*) ratio $L(\theta)/L(0.5)$. A lod score of 3.0 represents odds of ~1,000 to 1 that the observations are the result of linkage at the specified recombination frequency rather than independent segregation.

cies²⁹. To evaluate the sensitivity of linkage results to alternative values for these parameters, analyses were performed over a wide range of possible values. This range (0.55–0.95) included values within one standard error of the maximum penetrance estimates obtained from segregation analyses of the entire OOA sample as well as from pedigree 110 alone. Maximum penetrance values from the entire OOA sample increase to a value of 0.63 at age 30 or greater (Table 1). The values obtained for pedigree 110 when analysed separately from the other OOA families increased to a maximum penetrance of 0.85 for age 30 or greater (Table 1). The estimate of the trait allele frequency obtained from this family was 0.015.

Evaluation of the likelihood of linkage between *HRAS1* and the proposed affective disorder locus (as lod scores: see Table 2 for definition) provided strong support for linkage between the two loci (Table 2). A lod score of 3.340 at $\theta = 0$ for linkage between the proposed affective disorders locus and *HRAS1* was obtained at a maximum penetrance value of 0.63 based on segregation analyses of the total sample of 32 bipolar families. At this value of penetrance, the lod score continues to be above 3.0 for values of recombination frequency up to 0.07. Given these assumptions, the hypothesis of linkage between the two loci is favoured over the null hypothesis of non-linkage by odds of better than 1,000 to 1.

The lod score was 4.083 at $\theta = 0$ when the maximum penetrance value, 0.85, from segregation analyses of pedigree 110, was used. The lod score remained greater than 3.0 for all values of recombination frequency up to 0.25. If this estimate of penetrance is most accurate for pedigree 110, then odds favouring linkage as the explanation for the data reported here are >10,000 to 1 over the null hypothesis of non-linkage and chance. Odds favouring linkage at $\theta = 0$ remain high for all values of penetrance tested between 0.55 and 0.95 (data not shown). We conclude that linkage analysis supports the location of the proposed affective disorders locus in the region of chromosome 11 which includes the *HRAS1* gene.

To test this conclusion, linkage analyses were carried out between the *INS* locus, located ~3 map units from the *HRAS1* locus, and the affective disorders locus, using the same wide range of penetrance values included in the *HRAS1* analyses. The lod scores obtained provide support for a localization of

an affective disorders gene to the short arm of chromosome 11. The lod score was 1.75 at $\theta = 0.03$ (Table 3) using population based estimates of penetrance. Using penetrance values from pedigree 110 yielded a lod score of 2.63 at $\theta = 0.0$. Several factors could account for the somewhat lower values for *INS* compared to *HRAS1*. The most significant is that the affective disorder gene is probably closer to the *HRAS1* locus.

To assess the impact of variation in allele frequencies on the robustness of the findings, analyses were performed varying these estimates. For both *INS* and *HRAS1*, the allele frequencies used in the original analyses were compared to an arbitrary assumption of equal allele frequencies for all marker alleles. The lod scores increased slightly for *HRAS1* and decreased slightly for *INS* under these arbitrary assumptions (D.L.P., unpublished). In addition, analyses were performed incorporating alternative values for allele frequency of the affective-disorder locus. Allele frequencies of 0.010, 0.015, 0.019 and 0.023 were included. In all cases, analyses indicate that the conclusions we reach are not dependent on allele frequencies used as all estimates gave scores with only minor numerical differences. We conclude that the data provide strong support for linkage between the affective disorders locus and the *HRAS1* and *INS* loci.

Multipoint linkage analysis

To evaluate more accurately the relative position of the affective disorders gene and the *INS* and *HRAS1* loci, multipoint linkage analyses were performed using the computer program LINK-MAP³⁰. Multipoint analyses included information already known about linkage of marker loci by specifying the distances between them and then examining the likelihood of an unmaped locus being at various positions relative to the specified positions of the markers. Three-point analyses were performed to test for linkage of the proposed affective disorder locus to the *INS*–*HRAS1* region. Data on the relative map positions of *INS* and *HRAS1* were derived through analysis of segregation of these markers in the Venezuela Reference Pedigree and this OOA pedigree²² (D.S.G. *et al.*, in preparation). Results strengthen and are consistent with the pairwise findings (Fig. 2). The lod score exceeds 3.0 at all map positions within 30 cM of the *INS*–*HRAS1* region. The slight reduction in the lod score at *INS* (lod = 4.083) implies that the gene for bipolar affective disorder is less likely to be located exactly at the position of this marker. The lod score peaks (at 4.904) at *HRAS1*. These results indicate that the affective disorder gene is most likely to be tightly linked to the *HRAS1* locus on the short arm of chromosome 11.

Implications

It has been suggested that the complex nature of common disorders represents a serious barrier to the application of genetic linkage strategies³¹. For the first time the application of a generalized strategy for scanning the human genome has provided convincing evidence for the localization of a gene that

Table 3 The lod scores for pairwise linkage analyses between *INS* and a major affective disorders locus for two sets of penetrance estimates

Maximum penetrance	Recombination frequency (θ)						
	0.000	0.001	0.010	0.030	0.050	0.070	0.100
0.63	1.671	1.677	1.714	1.752	1.754	1.733	1.674
0.85	2.634	2.632	2.615	2.566	2.503	2.431	2.308

Analyses were performed as in Table 2 except that the following allele frequencies (derived from pedigree 110) for *INS* were used: allele 1 (haplotype A), 0.40; allele 2 (haplotype B), 0.15; allele 3 (haplotype C), 0.39 and allele 4 (haplotype D), 0.06. Data were relabelled to reduce computation time. Relabelling affects numerical results only by altering their specific dependency on marker allele frequency⁴³.

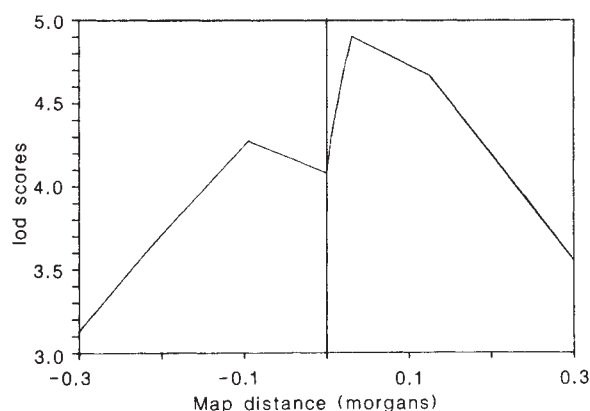


Fig. 2 Multipoint linkage analysis for bipolar affective disorder versus *INS* (at 0.0 morgans) and *HRAS1* (0.031 morgans).

Methods. Analyses using the computer program LINKMAP³⁰ were performed incorporating the gene frequency and age-specific penetrance estimates from segregation analyses of pedigree 110: $q = 0.015$; $f_{A-} = 0.18$, $f_{aa} = 0.0001$ for ages 15–19; $f_{A-} = 0.42$, $f_{aa} = 0.001$ for ages 20–24; $f_{A-} = 0.73$, $f_{aa} = 0.001$ for ages 25–29; $f_{A-} = 0.85$, $f_{aa} = 0.001$ for ages 30 and over. Quantities as in Table 1. The trait and marker data used were identical to those used in the pairwise analyses. However, one loop in the pedigree structure had to be broken to allow the program to run. Based on pairwise analyses, the loop was broken in the most conservative way.

is implicated in the aetiology of a common clinical disorder. These findings have broad implications for research in human genetics and psychiatry. The demonstration by a linkage strategy that a simple genetic mechanism can account for the transmission of bipolar affective disorders in pedigree 110 should provide an impetus to analogous research on other common clinical conditions. These illnesses include cancer, cardiovascular disease, diabetes and hypertension as well as other psychiatric disorders.

The results reported raise two major questions in psychiatric genetics. First, to what extent do the genetics of bipolar affective disorders in pedigree 110 reflect the genetics of affective disorders in other families and populations? Second, through what molecular mechanism does a chromosome 11 DNA sequence predispose members of pedigree 110 to bipolar affective disorders?

Genetic and clinical heterogeneity

The other families identified with affective disorders during the Amish study can be of great value in addressing the first of these questions. Data reported here on pedigree 110 show that a number of clinically variant forms of affective disorder (bipolar I, bipolar II, major depressive and schizoaffective) share a common chromosome 11 haplotype. Analysis of direct extensions of pedigree 110 and expansion of linkage studies to other Amish families identified with affective disorders could provide further data on the clinical spectrum associated with the chromosome 11 gene. It will be important to determine whether linkage with chromosome 11 markers is observed in multi-generational Amish families that exhibit forms of major affective disorders other than bipolar illness. If additional families that carry the chromosome 11 affective-disorders locus can be identified, then the existence of genes or environmental factors that modify the expression of penetrance of this locus could be effectively studied.

Although a great deal can be learned through extensions of the Amish study, it is also clear that linkage of chromosome 11 markers for predisposition to affective disorders must be tested in other family studies. A recent report describes cosegregation of thalassaemia and affective disorder in a non-Amish pedigree³². As the β -globin locus responsible for β -thalassaemia is located on the short arm of chromosome 11 and is closely

linked to the *HRAS1* and *INS* loci^{22,23} these findings would indicate that a chromosome 11 affective disorders gene may have been detected in at least one non-Amish family.

Heterogeneity is likely in a syndrome as common as bipolar disorder. A variety of samples of affective disorder families will be required to explore this issue. Given the difficulties of ascertainment of multi-generational families in the population at large, the affected sib pair method³³ is an attractive alternative to complete family studies. A sample of 100–200 affected sib pairs should rapidly determine the extent to which the chromosome 11 locus contributes to the aetiology of affective disorders in the general population.

Heterogeneity would be demonstrated if linkage could be found between a locus for predisposition to affective disorder and genetic markers located in other regions of the genome. Several investigations have suggested possible linkage to markers (colour blindness and the Xg locus) on the X chromosome. Because Xg and protan/deutan colour blindness loci are located at opposite ends of the X chromosome, the results of these studies conflicted. A number of further studies, using different samples, did not confirm these initial findings³⁴. Evidence suggesting linkage between a gene for affective illness and the chromosome 6 locus for human leukocyte antigen system (*HLA*) has been reported^{35,36}. Subsequent investigations have been unable to replicate these findings^{37,38}. The data from the Amish study of bipolar pedigrees conclusively exclude linkage to *HLA* in this population¹⁴.

Although genetic heterogeneity may explain some apparent contradictions among studies, use of different analytical approaches could also contribute to conflicting outcomes. In previous studies, positive findings were specific to narrow ranges of genetic model parameters. These parameters can significantly influence the outcome of linkage analyses²⁹. Thus, these studies do not provide convincing evidence for a possible locus involved in the aetiology of major affective disorders.

The findings reported here are robust over a wide range of genetic model parameters. The lod scores reported for linkage between *HRAS1* and the proposed affective disorder locus are larger than 3.00 for all values of maximum penetrance greater than 0.55 and exceed 4.00 for values of maximum penetrance from 0.85 to 0.95. Furthermore, the application of multipoint linkage analysis provides even stronger support than two-point analyses for the linkage of the proposed affective disorder locus to *INS* and *HRAS1*. Analysis of other RFLPs scattered throughout the genome in the original 51 individuals collected of pedigree 110 yielded no evidence for linkage to those markers³⁹. This further strengthens the chromosome 11 findings reported here.

Clinical implications

Although the genetics of affective disorders in the general population is likely to be more complex than in pedigree 110, it is clear that DNA-based diagnostic procedures have a future in psychiatry. The ability to identify individuals at risk by DNA genotype promises significant clinical benefits. Because not all individuals who inherit the gene will manifest the illness, the ability to identify those at risk by direct DNA analysis could permit the design of new studies to examine the interaction of genetic and environmental factors important for the onset of illness. Such prospective studies, although attractive theoretically, raise serious ethical and social considerations. Similar caution must be exercised in the direct use of predictive information in a clinical setting.

The second major issue raised by our findings is the precise identification of the physiological or biochemical mode of action of the gene responsible for bipolar disorders within pedigree 110. The clinical response of patients to therapeutic agents provides a framework for integration of insights from molecular biology once this gene is identified. As a number of expressed DNA sequences have been mapped to the short arm of chromo-

some 11, several possible candidate DNA sequences for the proposed affective disorders locus have already been identified. For example, the structural gene encoding tyrosine hydroxylase, mapped to the same segment of chromosome 11 as *INS* and *HRA1*⁴⁰, deserves consideration, because this enzyme catalyses an important step in the dopamine synthesis pathway. It has been suggested that defects in dopamine metabolism play an important role in the development of major mental disorders⁴¹.

Elucidation of the mechanism of action of the affective-disorders gene should lead to a better understanding of the aetiology of affective illness. It will be satisfying if insight into the molecular genetics of affective disorders leads to improved treatment for this serious and prevalent psychiatric condition.

This work was supported by grants from the National Institutes of Mental Health (MH28287 MH00508, MH28274, MH30929 and MH39239) and from NIH (GM27882). D.L.P.

Received 20 October; accepted 29 December 1986.

- Whybrow, P. C., Akiskal, H. S. & McKinney, W. T. *Mood Disorders: Toward a New Psychobiology* (Plenum, New York, 1984).
- Robins, L. N. et al. *Archs gen. Psychiat.* **41**, 949–958 (1984).
- Jefferson, J. W., Greist, J. H., Ackerman, D. L. & Carroll, J. A. *Lithium Encyclopedia for Clinical Practice* (American Psychiatric Press, Washington, 1986).
- Mindham, R. H. S. in *Handbook of Affective Disorders* (ed. Paykel, E. S.) 231–245 (Guilford, New York, 1984).
- Nies, A. & Robinson, D. S. in *Handbook of Affective Disorders* (ed. Paykel, E. S.) 246–261 (Guilford, New York, 1984).
- Bertelsen, A. in *Origin Prevention and Treatment of Affective Disorders* (eds Schou, M. & Stromgren, E.) 227–239 (Academic, London, 1979).
- Cadore, R. M. *J. Psychiat.* **135**, 463–466 (1978).
- Von Knorring, A., Cloninger, C. R., Bohman, M. & Sigvardsson, S. *Archs gen. Psychiat.* **40**, 943–950 (1983).
- Mendlewicz, J. & Rainer, J. D. *Nature* **268**, 327–329 (1977).
- Gershon, E. S. in *Psychiatry Update* (ed. Grinspoon, L.) 434–456 (American Psychiatric Press, Washington DC, 1983).
- Egeland, J. A. & Hostetter, A. M. *Am. J. Psychiat.* **140**, 56–61 (1983).
- Egeland, J. A. (ed.) *Descendants of Christian Fisher and other Amish and Mennonite Pioneer Families* (Johns Hopkins, Baltimore, 1972).
- Pauls, D. L. & Egeland, J. A. *Proc. natn. Acad. Sci. U.S.A.* (submitted).
- Kidd, K. K. et al. *Am. J. Psychiatry* **141**, 1042–1048 (1984).
- Housman, D. E. & Gusella, J. F. in *Psychobiology and Psychopathology Vol. 1: Genetic Research Strategies in Psychobiology and Psychiatry* (eds Gershon, E. S., Matthysse, S., Breakefield, X. O. & Ciaranello, R. D.) 17–24 (Boxwood, Pacific Grove, 1981).
- Egeland, J. A. in *Catalogue of Cell Lines: NIH Publication 85–2011*, 563–567 (National Institutes of Health, Washington DC, 1985).
- Gerhard, D. S. et al. *Am. J. hum. Genet.* **36**, 3S (1984).
- Kidd, K. K., Gerhard, D. S., Kidd, J. R., Housman, D. & Egeland, J. A. *Clin. Neuropharmac.* **7**, 198–199 (1984).
- Endicott, J. & Spitzer, R. L. *Archs gen. Psychiat.* **35**, 837–844 (1978).
- Spitzer, R. L., Endicott, J. & Robins, E. *Archs gen. Psychiat.* **35**, 773–782 (1978).

was supported by an RSDA award from NIMH and D.S.G. by a postdoctoral fellowship from Damon Runyon-Walter Winchell Foundation DRG-665. We wish to acknowledge the contribution of Drs Jean Endicott, David R. Offord and John J. Schwab to the clinical aspects of this study. We thank Dr Arthur E. Greene and the Institute for Medical Research (now the Coriell Institute) of Camden, New Jersey, for the development of transformed lymphocyte cell lines from pedigree 110 and Judy Kidd for assistance in the preparation of some of the DNAs. We thank Dr G. Bell for the probe phins310, Dr R. Weinberg for pEJ 6.6, and Drs A. J. Pakstis and P. L. Kramer for assistance in the multipoint linkage analyses. We thank Jen Cunningham for help in preparation of the manuscript; Ruth Housman, Karen Kabnick and Rachel Kent for helpful discussions and suggestions; and Sam Barondes, Jean Endicott, David Offord, Herb Pardes and Nancy Wexler for valuable critiques of the final text.

- Hostetter, A. M., Egeland, J. A. & Endicott, J. *Am. J. Psychiat.* **140**, 62–66 (1983).
- Gerhard, D. S., Kidd, K. K., Housman, D. E., Gusella, J. & Kidd, J. R. *ICSU Short Rep.* **1**, 172–173 (1984).
- White, R. et al. *Nature* **313**, 101–105 (1985).
- Bell, G. I., Selby, J. J. & Rutter, W. J. *Nature* **295**, 31–35 (1982).
- Capon, D. J., Chen, E. Y., Levinson, A. D., Seeborg, P. H. & Goeddel, D. V. *Nature* **302**, 33–37 (1983).
- Chakravarti, A., Elbein, F. C. & Permutt, M. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1045–1049 (1986).
- Ott, J. *Am. J. Hum. Genet.* **28**, 528–529 (1976).
- Hodge, S., Morton, L., Tideman, S., Kidd, K. K. & Spence, M. A. *Am. J. hum. Genet.* **31**, 761–762 (1979).
- Kruger, S., Turner, W. & Kidd, K. *Biol. Psychiat.* **17**, 1081–1099 (1982).
- Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Am. J. hum. Genet.* **37**, 482–498 (1985).
- Sturt, E. & McGuffin, P. *Psychol. Med.* **15**, 455–462 (1985).
- Joffe, R. T., Harvath, Z. & Tarvydas, I. *Am. J. Psychiat.* **143**, 933 (1986).
- Suarez, B. K., Rice, J. & Reich, T. *Ann. hum. Genet.* **42**, 87–94 (1978).
- Gershon, E. S. & Bunney, W. J. *Psychiat. Res.* **13**, 99–117 (1976).
- Smeraldi, E., Neri, F., Melica, A. & Scorza-Smeraldi, R. *Tissue Antigens* **12**, 270–274 (1978).
- Weikamp, L., Stancer, H., Persad, E., Flood, C. & Gutterormsen, C. *New Engl. J. Med.* **305**, 1301–1306 (1981).
- Goldin, L., Clerget-Darpoux, F. & Gershon, E. *Psychiat. Res.* **7**, 29–45 (1982).
- Suarez, B. K. & Reich, T. *Archs gen. Psychiat.* **41**, 22–27 (1984).
- Kidd, J. R. et al. *J. Psychiat. Res.* (in the press).
- Craig, S. P., Buckle, V. J., Craig, I. W., Lamouroux, A. & Mallet, J. *Cytogenet. & Cell Genet.* **40**, 610 (1985).
- Schatzberg, A. S., Rothschild, A. J., Langlais, P. J., Bird, E. D. & Cole, J. O. *J. Psychiat. Res.* **19**, 57–64 (1985).
- Lalouel, J. M., Rao, D. C., Morton, N. E. & Elston, R. C. *Am. J. hum. Genet.* **35**, 816–826 (1983).
- Braverman, M. S. *Computers Biomed. Res.* **18**, 24–36 (1985).
- Gerhard, D. S., Kidd, K. K., Kidd, J. R., Egeland, J. A. & Housman, D. E. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7875–7879 (1984).

LETTERS TO NATURE

Shocked bipolar outflow from the evolved star OH231.8+4.2

Bo Reipurth

European Southern Observatory, Casilla 19001, Santiago 19, Chile

Low- and intermediate-mass stars lose significant fractions of their mass during the asymptotic-giant-branch evolution, creating massive circumstellar shells. The deeply enshrouded OH infrared stars are presumably at the end of asymptotic-giant-branch evolution. I report here the discovery of shocked bipolar bubbles expanding supersonically at right angles to a dense dust disk around the red-giant OH infrared star OH231.8+4.2. The relative radial velocity between the northern and southern bubble fronts is over 200 km s^{-1} , and with a total physical extent of 0.42 pc and a known inclination, this gives a dynamical age of about 1,500 yr. Herbig-Haro objects are observed at the front of the bubbles, and are probably formed as instabilities in the bow shocks, where two collimated flows ram into the ambient medium. This new phenomenon represents a brief phase of late stellar evolution, immediately preceding the formation of a planetary nebula.

OH231.8+4.2 (also known as OH0739–14) is a strong OH infrared source, surrounded by an elongated near-infrared reflection nebula of $\sim 9 \text{ arc s}$, which coincides with a faint optical nebulosity extended along the same direction². Optical spectroscopy³ indicates a spectral type of the hidden star of M9. Monitoring of infrared luminosity⁴ and OH-maser intensity⁵ shows a regular periodicity with a period of ~ 650 –680 days. Mapping of the 1,667-MHz OH maser emission with high spatial and velocity resolution shows bipolar large-scale velocity gradients along the optical-infrared axis^{5,6}. Recently, Cohen *et al.*⁷ performed long-slit spectroscopy along the axis of the optical nebulosity, and discovered Herbig-Haro (HH) like high-velocity blue- and redshifted shock-excited regions outside the nebula.

Direct charge-coupled device (CCD) imaging was carried out at the Danish 1.5-m telescope at the European Southern Observatory (ESO), La Silla, with a GEC chip having low read-out noise and a pixel size of $22 \mu\text{m}$, corresponding to 0.35 arc min on the sky. Figure 1a shows a deep H α /[N II] interference filter image of OH231.8+4.2. The immediately striking features are the two bubble-shaped lobes emanating from the central object, two collimated streamers and an obscuring ridge separating the southern and northern parts. With an axial extent of 30 and 20 arc s for the southern and northern bubbles, respectively, these are very large objects; at a distance of 1.2 kpc ⁸

Table 1 Infrared observations

Wavelength (μm)	Magnitude	Flux (Jy)
1.25	8.19(0.02)	0.85(0.01)
1.65	6.72(0.01)	2.18(0.03)
2.2	5.56(0.02)	3.88(0.06)
3.8	3.28(0.02)	12.3(0.22)
4.8	2.19(0.04)	20.6(0.68)
12		19.0(1.5)
25		226(18)
60		548(44)
100		292(47)

1.25–4.8 μm : ESO 1-m telescope data.12–100 μm : IRAS data.

and with an inclination of 47° to the plane of the sky (Tielens *et al.*, in preparation) these correspond to 0.25 pc and 0.17 pc (50,000 and 35,000 AU) respectively. The HH objects recently discovered by long-slit spectroscopy are located at distances from the central source corresponding to the front edges of the bubbles. Figure 1b shows a similar exposure through a slightly redder filter, excluding the $H\alpha$ and $[\text{N II}]$ lines. Only reflected light and fore- and background stars are visible, and all traces of the bubble structures have vanished, testifying to their emission-line nature. Another similar exposure through a $[\text{S II}]$ filter, shows the basic features of the bubbles, but only at very low light levels. The morphology and evolution of this system is discussed later.

Long-slit spectra have been obtained with the ESO 2.2-m and 3.6-m telescopes, with Boller and Chivens spectrograph and CCD-detector, with the slit placed along the symmetry axis. The spectra confirm the finding of Cohen *et al.*⁷ that the $[\text{N II}]$ lines are of extreme intensity, and that the northern shock region is blueshifted and the southern redshifted. Based on the five lines $H\alpha$, $[\text{N II}]$ 6,548 Å, $[\text{N II}]$ 6,584 Å, $[\text{S II}]$ 6,717 Å and $[\text{S II}]$ 6,731 Å the heliocentric radial velocity of the northern shock region was measured to be -42 km s^{-1} , and for the southern region $+180 \text{ km s}^{-1}$, both with a standard deviation of 27 km s^{-1} . In the local standard of rest this corresponds to -60 km s^{-1} and $+162 \text{ km s}^{-1}$, respectively. Maser emission has been mapped with high spatial and velocity resolution⁶, showing that maser emission is almost symmetrically extended about 10 arc s along the polar axis, with $-16 \text{ km s}^{-1} < v_{\text{lsr}} < 81 \text{ km s}^{-1}$, and a velocity at the source location of $+20 \text{ km s}^{-1}$. The radial velocity of the northern shock front relative to the central maser velocity is lower than for the southern front (80 compared with 140 km s^{-1}), which possibly accounts for its less developed state. From the known inclination of the system, one finds space velocities of 110 km s^{-1} (N) and 190 km s^{-1} (S). With dimensions of 0.17 and 0.25 pc for the northern and southern bubbles, respectively, this suggests dynamical ages of $\sim 1,500$ and $\sim 1,300$ yr, which are identical within the uncertainties.

Near-infrared photometry was carried out with the ESO 1-m telescope, and is listed in Table 1, together with IRAS (infrared astronomical satellite) fluxes. These near-infrared data are slightly brighter than the brightest observations of the extended data set of Feast *et al.*⁴, so the present data were taken very near to maximum light. The present observation at JD 2446478 offers a long baseline to the early infrared data, and assuming conservatively that it falls within a half-year window around the true maximum of the lightcurve, and assuming another maximum at JD 2444403 in ref. 4, a period of 692 ± 30 days is suggested. The near-infrared variability is considerable, the luminosity between 1.25 and 2.2 μm for the two data sets in Fig. 2 varies from 34 to 96 $L_\odot$, close to a factor three. The energy distribution of OH231.8+4.2 from the data in Table 1 plus a 400 μm observation from Sopka *et al.*⁸ is shown in Fig. 2. The luminosity of the data in Table 1 between 1.25 and 4.8 μm is about 420 $L_\odot$, and the data in Fig. 2 give a luminosity between

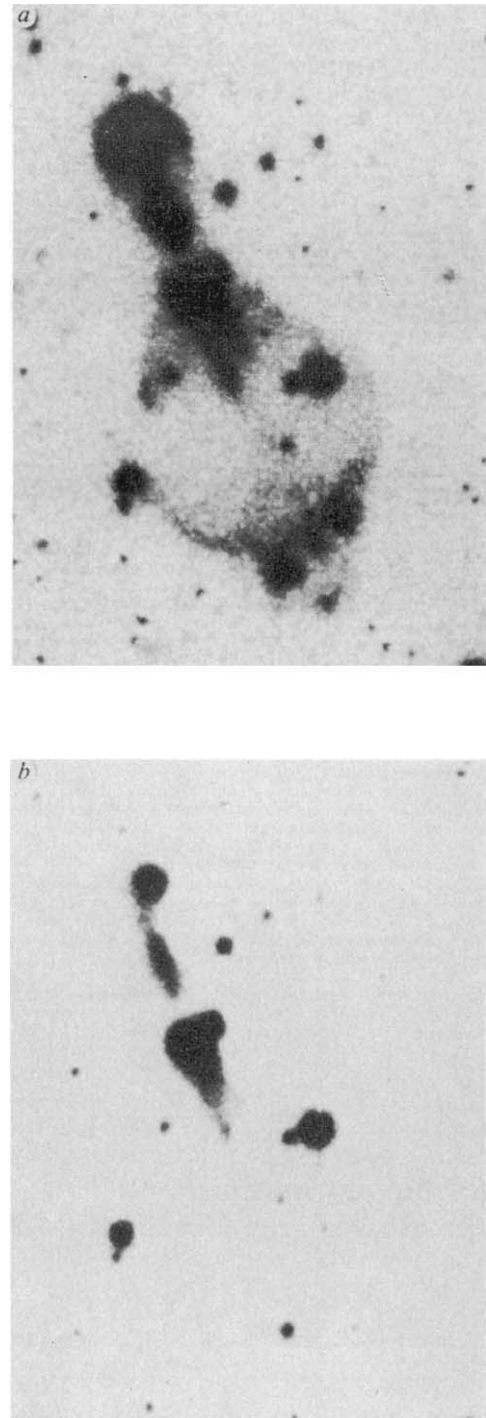


Fig. 1 CCD images of OH231.8+4.2 obtained with the Danish 1.5-m telescope at ESO, La Silla, through different interference filters, on 6 March 1986, both 60-min exposures. *a*, $H\alpha/[\text{N II}]$ image centred on 6,565 Å and *b*, continuum image centred on 6,646 Å, both with FWHM (full width at half maximum) bandwidth of 78 Å. The images demonstrate that the bubble structures are purely line-emission objects. The total extent of the object is 50 arc s. North is up and east is left.

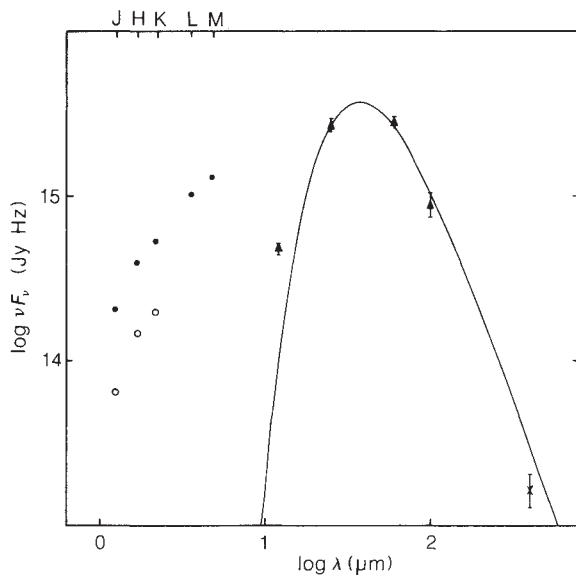


Fig. 2 The energy distribution of OH231.8+4.2. Filled circles are observations from this paper through a 15-arc s diaphragm, close to maximum light. Open circles are from ref. 4, through a 36-arc s diaphragm, on JD 2444673, close to minimum light. JHKLM denote the 1.25 to 4.8 μm photometric bands listed in Table 1. Triangles are fluxes from the IRAS Point Source Catalogue. The IRAS fluxes as plotted are approximate, because (1), the object is listed as variable; (2), the object is listed as a small extended source; and (3), no colour-correction factors were applied. The cross is a 400- μm measurement by Sopka *et al.*⁸. Error bars of the near-infrared measurements are less than the extent of the circles. The solid line represents a black body at a temperature of 100 K.



Fig. 3 A schematic representation of OH231.8+4.2, based on a contour plot of the $\text{H}\alpha/[\text{N II}]$ CCD image in Fig. 1a. A disk surrounds the central star, which is invisible along our line of sight. Collimated bipolar outflows emerge above and below the disk. The surface brightness of the luminous southern lobe decreases towards the star because of extinction in the disk. A similar, but much less pronounced effect is seen in the narrow northern lobe, proving that the dust envelope is indeed primarily a highly flattened disk structure and not a near-spherical envelope. This is also indicated by the 'horns' or two-fingered appearance of the second brightest contour in the southern lobe, because Morris⁹ has shown that when the dust density falls off rapidly with increasing angle from the equatorial plane, such a geometry occurs. Shocked bubbles are formed by collimated flows emerging from the central source and ramming into the ambient medium. Contour levels were chosen to outline important features and are in arbitrary units 285, 330, 490, 800, with a background level of about 265. The HH knots in the southern bow shock are inserted as dotted contours at a level of 305. Stars are shown as black circles, and cosmic rays (seen in Fig. 1a) have been erased.

12 and 400 μm of about 2,700 $L_{\odot}$. Much of this large luminosity is from dust heated to ~ 100 K.

Part of this dust may be visible as the dark obscuring lane separating the northern and southern lobes in the CCD images. This is evidently a direct optical detection of a circumstellar disk. The star must be situated inside the disk, buried at the very base of the narrow northern lobe, because the northern flow axis is inclined towards us. Therefore, the brightest spot seen in all CCD images as well as in the infrared maps of Allen *et al.*¹ is not the location of the star, but simply a part of a reflection nebula, which fades to the south because of increasing distance from the illuminating source, and fades to the north because of increasing extinction in the disk. The absolute position of the star is known from the OH observations, but cannot be compared with the CCD images, for which astrometric positions have not been obtained. However, the infrared position was determined by Allen *et al.*¹, and does not coincide with the stellar position derived from the maser observations, but is in fact displaced 2 arc s southwards along the flow-axis. In the 2 μm map of Allen *et al.*¹, the intensity decreases towards the star over this distance by a factor of ten, which corresponds to an increase in extinction of 2.5 mag at K, if the underlying emission is constant, but the underlying emission certainly increases strongly towards the star, and the above value is therefore very much a lower limit. Assuming a normal reddening law (not necessarily a good assumption in a circumstellar disk) this then corresponds to an increase in visual extinction along the disk radius ≥ 20 mag.

Figure 3 shows the interpretation discussed above. The contours are drawn from the $\text{H}\alpha/[\text{N II}]$ image shown in Fig. 1a, and all field stars are shown as filled circles. The distance from the centre of the disk to the centre of the brightest contour in the southern lobe is about 3 arc s (larger than in the 2- μm map, as expected from the higher extinction in the optical). With the disk inclined about 47° to the line of sight and at a distance of 1.2 kpc, this suggests a lower limit to the disk radius of

$\sim 5,000$ AU. This compares with the radius of $\geq 6,000$ AU determined from the ring of OH masers surrounding the embedded star⁵.

The shocked bubbles around OH231.8+4.2 are obviously formed by a collimated outflow that rams into the ambient medium. This medium must be denser than the general interstellar medium for the two polar streams to produce the large and bright shock structures. Rather, the present collimated high-velocity wind is probably encountering a previous, slow, low-density outflow, producing the shocked emission at the interface zone. Indeed, from the clear limb-brightening of the southern lobe it is obvious that it is not filled with a uniform, luminous shocked gas, but that the shocks occur in a thin layer at the bubble boundary. The bow-shock of the northern lobe appears rather smooth within the present resolution, whereas the southern bow shock shows three condensations, with another isolated knot below them. These HH objects are apparently formed as a dynamical instability or a cooling instability in the bow-shock. The northern bubble is obviously less developed than the southern. This could either be because of a density-

gradient in the older, low-density wind, through which the present flow rams, or an asymmetry in the collimating disk. It might also be an intrinsic property of the bipolar outflow itself. Very high-resolution molecular observations may resolve this. The reason that the most dense circumstellar material is located in a disk is perhaps related to the central star being a binary star^{7,10}. The flow itself may be driven by radiation pressure on dust grains, a mechanism suggested to drive mass loss from highly evolved red-giant stars. The fact that the momentum flux in the flow from OH231.8+4.2 is roughly an order of magnitude larger than the momentum flux in the present stellar radiation^{11,12} does not necessarily violate this hypothesis, because the star is clearly in a rapidly developing phase, as also suggested by the change from a slow to a fast wind^{13,14}.

The next stage for OH231.8+4.2 is presumably the formation of a planetary nebula. At the end of the heavy mass-loss phase, the ultraviolet radiation from the then naked stellar core will ionize the ambient medium. Unless the massive disk has been completely dispersed at that time (which seems unlikely), the shadow effect of the remaining disk material will cause a bipolar geometry of the planetary nebula. A radio continuum search with the VLA (very large array) shows that OH231.8+4.2 has not yet developed a directly observable ionized core¹⁵.

This project was supported by ESO and the Danish Board for Astronomical Research. I thank M. Morris and D. Reimers for valuable comments.

Received 1 September; accepted 25 November 1986.

- Allen, D. A., Barton, J. R., Gillingham, P. R. & Phillips, B. A. *Mon. Not. R. astr. Soc.* **190**, 531-536 (1980).
- Cohen, J. G. & Frogel, J. A. *Astrophys. J.* **211**, 178-180 (1977).
- Cohen, M. *Publ. Astron. Soc. Pacific* **93**, 288-290 (1981).
- Feast, M. W. *et al. Mon. Not. R. astr. Soc.* **203**, 1207-1211 (1983).
- Bowers, P. F. & Morris, M. *Astrophys. J.* **276**, 646-652 (1984).
- Morris, M., Bowers, P. F. & Turner, B. E. *Astrophys. J.* **259**, 625-633 (1982).
- Cohen, M., Dopita, M. A., Schwartz, R. D. & Tielens, A. G. G. M. *Astrophys. J.* **297**, 702-709 (1985).
- Sopka, R. *et al. Astrophys. J.* **294**, 242-255 (1985).
- Morris, M. *Astrophys. J.* **249**, 572-585 (1981).
- Zuckerman, B. & Aller, L. H. *Astrophys. J.* **301**, 772-789 (1986).
- Zuckerman, B. & Dyck, H. M. *Astrophys. J.* **304**, 394-400 (1986).
- Knapp, G. R. *Astrophys. J.* **311**, 731-741 (1986).
- Jura, M. & Morris, M. *Astrophys. J.* **292**, 487-493 (1985).
- Jura, M. *Astrophys. J.* **286**, 630-632 (1984).
- Spergel, D. N., Giuliani, J. L. & Knapp, G. R. *Astrophys. J.* **275**, 330-341 (1983).

Neutrino oscillations in the Earth suggest a terrestrial test of solution to solar neutrino problem

Arnon Dar & Ady Mann

Departments of Physics and Space Research Institute,
Technion-Israel Institute of Technology, Haifa, Israel

It is generally believed that a complete spectroscopy of solar neutrinos, which is far from being achieved in the near future, is required in order to verify the Mikheyev-Smirnov-Wolfenstein (MSW) solution^{1,2} of the solar neutrino problem^{3,4} and to unfold the values of neutrino masses and mixings that nature has selected. However, as we discuss here, the MSW solution^{1,2} of the solar neutrino problem^{3,4} can be tested with present equipment—the 'Bethe solution'⁵ yields sizeable oscillations of atmospheric neutrinos in the Earth which can be detected with the massive underground proton decay detectors, while the 'Rosen-Gelb' additional solutions⁶ yield diurnal and seasonal modulations of the solar neutrino flux which perhaps can be detected by the radiochemical Cl and Ga detectors. Moreover, neutrino oscillations in Earth may modify the values of the oscillation parameters which can solve the solar neutrino problem and help determine their values.

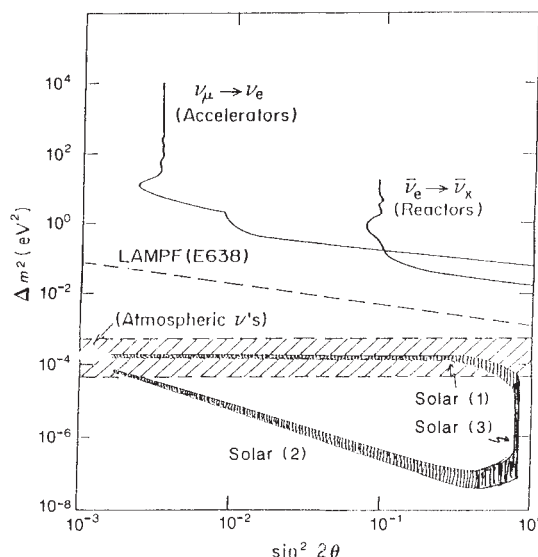


Fig. 1. Summary of experimental results on Δm^2 and $\sin^2 2\theta$. The areas to the right hand side of the full lines are excluded by neutrino oscillation experiments at accelerators and reactors. The area above the dashed line will be tested within a few years at Los Alamos Meson Production Facility (LAMPF). The shaded area can be tested by searching oscillations of atmospheric neutrinos with $0.1 \leq E_\nu \leq \infty$ GeV. The narrow black strips indicate the values that can solve the solar neutrino problem^{1,5-7}.

The 'solar neutrino problem' consists of the well-known discrepancy between the expected production rate of ^{37}Ar atoms by solar neutrinos through the reaction $\nu_e + ^{37}\text{Cl} \rightarrow e^- + ^{37}\text{Ar}$, and the production rate that has been observed by R. Davis and collaborators since 1970. Whereas the standard solar model with the best available input physics predicts a neutrino capture rate of $7.6 \pm 3.3(3\sigma)$ SNU (refs 3, 4) the observed capture rate is only 2.1 ± 0.3 SNU (refs 3, 4) (1 SNU = 10^{-36} captures per ^{37}Cl atom per second). In spite of many attempts no simple explanation for this discrepancy has been found³. However, recently Mikheyev and Smirnov¹ offered a plausible solution to the solar neutrino problem. They found by numerical integration of the neutrino propagation equations² in the Sun that thermonuclear ν_e s which are produced in the core of the Sun can convert into ν_μ s before emerging from the Sun. These escape detection in the chlorine solar neutrino detector of R. Davis because their energy $E_\nu \leq 14$ MeV, is below the threshold energy for the reaction $\nu_\mu + ^{37}\text{Cl} \rightarrow \mu^- + ^{37}\text{Ar}$.

The discovery of Mikheyev and Smirnov has been confirmed by independent calculations by Bethe⁵, by Rosen and Gelb⁶ and by Dar *et al.*⁷. The detailed analysis^{1,5-7} shows that neutrino oscillations in the Sun can solve the solar neutrino problem only if nature has selected specific combinations of neutrino masses and mixing angle. If ν_e and ν_μ , the neutrino flavour eigenstates which are produced in the weak interaction, are not eigenstates of the free hamiltonian but a linear combination of the mass eigenstates ν_1 and ν_2 with masses m_1 and m_2 respectively, that is $\nu_e = \cos \theta \nu_1 + \sin \theta \nu_2$ and $\nu_\mu = \cos \theta \nu_2 - \sin \theta \nu_1$, then the solar neutrino problem can occur because of neutrino oscillations in the Sun if the masses and mixing satisfy either^{1,5-7}

- (1) the Bethe condition⁵: $\Delta m^2 \cos 2\theta \sim 1.2 \times 10^{-4} \text{ eV}^2$ and $\tan^2 2\theta > 10^{-3}$, or
- (2) the Rosen-Gelb condition⁶: $\Delta m^2 \sin^2 2\theta / \cos 2\theta \sim 4 \times 10^{-8} \text{ eV}^2$ and $\tan^2 2\theta > 10^{-3}$, or
- (3) the condition⁷: $1 \times 10^{-4} \leq \Delta m^2 \leq 4 \times 10^{-8} \text{ eV}^2$ and $\sin^2 2\theta \sim 2/3$,

where $\Delta m^2 = m_2^2 - m_1^2$. Conditions (1)–(3) are indicated in Fig. 1.

If nature has selected neutrino masses and mixing angles that satisfy condition (1) then the pp neutrinos do not convert into

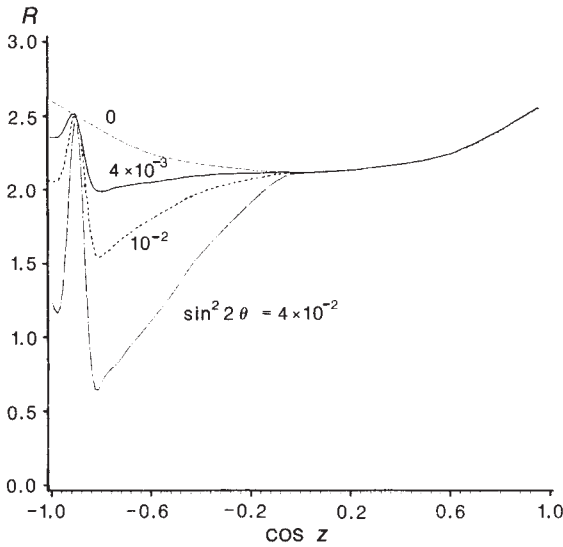


Fig. 2 The expected ratios between atmospheric ν_μ s and atmospheric ν_e s as function of zenith angles, for $E_\nu \sim 300$ MeV, $\Delta m^2 = 1.2 \times 10^{-4} \text{ eV}^2$, and for various values of $\sin^2 2\theta$.

ν_μ s in the Sun and the production rate of ${}^{71}\text{Ge}$ in ${}^{71}\text{Ga}$ should be that predicted by the standard solar model. In this case verification of the MSW solution of the solar neutrino problem in Cl cannot come from the Ga experiments⁸⁻¹⁰. It can, however, come from detecting sizeable oscillations of atmospheric neutrinos with $E_\nu \sim 300$ MeV in Earth, using the current massive underground proton decay detectors as neutrino telescopes (or from measuring the ratio between charge exchange reactions induced by solar ν_e s and the elastic scattering of both solar ν_e s and solar ν_μ s with (future) second generation proton decay detectors such as Kamiokande II, IMB II, and Icarus).

To see that let us consider the propagation of ν_e s and ν_μ s through Earth. In terms of Pauli's spin matrices σ_i ⁷

$$\frac{d}{dx} \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = i \frac{\pi}{l_v} (\alpha \sigma_1 + \gamma \sigma_3) \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix}, \quad (1)$$

with $\alpha = \sin 2\theta$ and $\gamma = l_v/l - \cos 2\theta$ where,

$$l_v = 4\pi E_\nu / \Delta m^2 = 2.48 \times E (\text{MeV}) / \Delta m^2 (\text{eV}^2) \text{ cm} \quad (2)$$

is the oscillation length in vacuum and⁵⁻⁷

$$l = \sqrt{2} \pi / N_e G_F = 1.64 \times 10^9 \text{ cm} / \rho_e \quad (3)$$

is the refraction length in matter, G_F is the Fermi coupling constant and ρ_e is the electron density divided by Avogadro's number; $\rho_e = N_e (\text{cm}^{-3}) / 6 \times 10^{23}$.

In a medium of constant density, equation (1) can be solved analytically by simple exponentiation. Using the algebra of the Pauli spin matrices one obtains

$$\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix}_x = [\cos(\pi x / l_m) + i(l_m / l_v) \sin(\pi x / l_m)(\alpha \sigma_1 + \gamma \sigma_3)] \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix}_0 \quad (4)$$

Note in particular that the probability of neutrino flavour flip is given by²

$$P(\nu_e \leftrightarrow \nu_\mu, x) = (l_m / l_v)^2 \sin^2 2\theta \cdot \sin^2(\pi x / l_m), \quad (5)$$

$$l_m = l_v / (\sin^2 2\theta + (\cos 2\theta - l_v / l)^2)^{1/2}, \quad (6)$$

where l_m is the oscillation length in matter. $P(\nu_e \leftrightarrow \nu_\mu)$ as a function of l_v , that is as a function of $E_\nu / \Delta m^2$, has a resonance shape around $l_v = l \cdot \cos 2\theta$ whose height is $\sin^2(\pi x \sin 2\theta / l_v)$ and its full width at half the height is $\Delta l_v = 2l \cdot \sin 2\theta$.

For mixing angles which are not too small, equations (4)–(6) describe well⁷ the propagation within the mantle and the core of Earth if one uses the average densities there instead of the

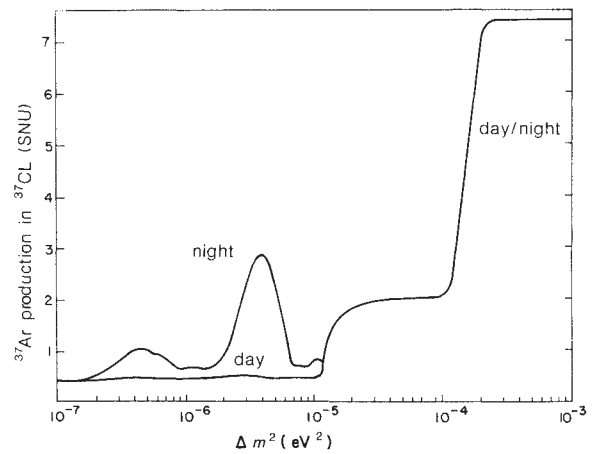


Fig. 3 The production rate of ${}^{37}\text{Ar}$ in ${}^{37}\text{Cl}$ at latitude 43° N as function of Δm^2 during day-time and during night-time, by solar neutrinos, for $\sin^2 2\theta = 0.10$.

exact density distributions¹¹. In the general case when the electron density changes along the neutrino trajectory equation (4) can be used to propagate the neutrinos in small steps⁷. We have used equation (4) and the density distribution of Earth from ref. 11 for calculating $P(\nu_e \leftrightarrow \nu_\mu)$, the probability of a neutrino flavour flip after crossing Earth. Our calculations show that a significant probability of neutrino flavour flip develops for $\sin^2 2\theta \geq 10^{-2}$ and zenith angles z satisfying $\cos z \leq -0.25$, for neutrinos which satisfy the resonance condition in Earth: the electron density of Earth changes from $\rho_e \sim 1.3$ near its surface to about $\rho_e \sim 6.0$ near its centre¹¹. Thus, the resonance condition $l_v = l \cdot \cos 2\theta$ in Earth is satisfied by neutrinos with $10^6 \leq E / \Delta m^2 \leq 5 \times 10^6 \text{ MeV/eV}^2$, that is by neutrinos with $120 \leq E \leq 600 \text{ MeV}$ if condition (1), $\Delta m^2 = 1.2 \times 10^{-4} \text{ eV}^2$, is satisfied.

The expected ratio between ν_μ s and ν_e s that reach the underground detector from below at zenith angle $z \geq \pi/2$ can be related to their ratio when they enter the Earth from the atmosphere at zenith angle $\pi - z$ on the far side of the Earth:

$$R \equiv (N_{\nu_\mu} / N_{\nu_e})_z = [(1 - P) \cdot R + P] / [P \cdot R + (1 - P)]_{\pi - z} \quad (7)$$

However, because geomagnetic and solar wind effects on the ratio ν_e / ν_μ are quite negligible, equation (7) can be tested with experimental data obtained at the same site and the experimental results at different sites can be combined to increase statistics.

In Fig. 2 we plotted R as function of zenith angle for atmospheric neutrinos with $E_\nu = 300$ MeV for representative values of $\sin^2 2\theta$ (4×10^{-2} , 1×10^{-2} , 4×10^{-3} and 0). The ratio ν_μ / ν_e for downgoing atmospheric neutrinos was taken from the calculations of Dar¹² which agree well both with experiments and with other independent calculations¹³. As can be seen from Fig. 2 the effect of neutrino oscillations on R for $\Delta m^2 = 1.2 \times 10^{-4} \text{ eV}^2$ is quite large for values of $\sin^2 2\theta$ which are not too small. We believe that such an effect, if indeed present, can be detected by the massive underground proton decay detectors because (i) the observed rate of interactions of atmospheric neutrinos in the proton decay detectors peaks around 300 MeV, which is the resonance energy in the Earth mantle of neutrinos that satisfy condition (1), (ii) the ratio N_{ν_μ} / N_{ν_e} is known¹⁴ to a precision of about 5%, (iii) the 'contamination' of the interactions of ν_e s and ν_μ s by antineutrinos is a small effect. (The fluxes of atmospheric antineutrinos are smaller and their cross sections are much smaller than those of neutrinos.) Therefore we suggest that the data on interactions of atmospheric neutrinos in massive underground proton decay detectors will be re-analysed to see whether there is any evidence for the 'Bethe solution'⁵.

If nature has selected neutrino masses and mixing angles that satisfy condition (2), or (3), then the rate of production of ${}^{71}\text{Ge}$ in ${}^{71}\text{Ga}$ decreases vastly from 126 SNU (73 SNU due to pp

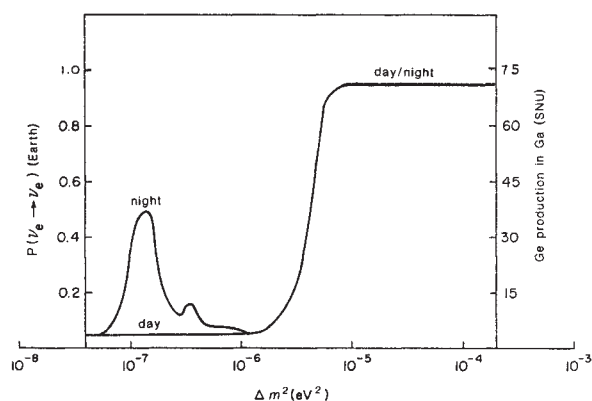


Fig. 4 The probability that pp neutrinos from the Sun reach a terrestrial detector at latitude 43° N as electron neutrinos, during day-time and during night-time, as function of Δm^2 , for $\sin^2 2\theta = 0.1$, and the corresponding rate of ^{71}Ge production in ^{71}Ga .

neutrinos) for $\Delta m > 10^{-5} \text{ eV}^2$ to practically zero for $\Delta m 10^{-6} < \text{eV}^2$, as shown in Fig. 4 by the 'day' line. But even if the Ga experiments do detect a signal much smaller than expected from the standard solar model and consistent with the MSW effect, it will neither verify the MSW effect nor will it establish the solar origin of the signals in the Cl or the Ga experiments. The expected capture rate for the decay hypothesis is indistinguishable⁷ from the expectation based upon the MSW effect with parameters satisfying condition (2).

But the estimated effects of the MSW solution on the Cl and the Ga experiments neglected^{1,5-7} neutrino oscillations in the Earth. At night solar neutrinos cross the Earth before reaching the detector. Passage through the Earth induces $\nu_e \leftrightarrow \nu_\mu$ transitions. Because the transition probability depends on the path in Earth, that is, on the position of the Sun relative to the detector, then neutrino oscillations in Earth may induce both diurnal and seasonal modulations of the production rates in radiochemical detectors and changes in the rate and in the spectrum of recoiling electrons in electron-recoil detectors.

The zenith angle of the Sun, z , as function of time at a geographical latitude ϕ ($\sim 43^\circ$ N for Homestake, Mont Blanc, Frejus, Grand Sasso and Baksan) is given by

$$\cos z = \sin \phi \sin \delta - \cos \phi \cos \delta \cos (2\pi h/23.934), \quad (8)$$

where h is the day time (in hours), 23.934 h is the mean sidereal day and δ is the Sun's declination as a function of time. $\sin \delta \approx -\sin (23.44^\circ) \cos (2\pi d/365.242)$, where 23.44° is the angle at which the ecliptic is declined to the plane of the celestial equator, 365.242 is the length in days of a tropical year and d is the time in the year (in days).

To estimate the diurnal effect we used equation (8) to perform the time averaging of $P(\nu_e \leftrightarrow \nu_\mu)$ which was calculated using equation (4) and the density distribution of Earth from ref. 11. The diurnal effect at latitude 43° N is demonstrated in Figs 3, 4 where we plotted the expected rates of ^{37}Ar production in ^{37}Cl , and of ^{71}Ge production in ^{71}Ga because of pp neutrinos, respectively, during day and night, as a function of Δm^2 for $\sin^2 2\theta = 0.1$. As can be seen from Figs 3, 4, if solar neutrinos are converted in the Sun into ν_μ s then the Earth may transform a large fraction of them, during night-time, back into ν_e s and the day-night difference in the radiochemical detectors may be quite large, (depending on the specific values of masses and mixing angles that nature has selected). Unfortunately, the ^{37}Ar in the ^{37}Cl experiment was extracted after exposures between 35 and 80 days (ref. 15 and R. Davis Jr, personal communication). To reveal a day-night modulation one will have to extract continuously the Ar atoms produced in Cl (the Ge atoms produced in Ga) into separate day and night containers and count

them separately, which apparently can be done with reasonable efficiency (R. Davis Jr, personal communication). However, a day-night difference, if it is sufficiently large should reveal itself also in a summer-winter difference. At 43° N the ratio between the average time that the Sun is below the horizon in the winter (1 October–31 March) and in the summer (1 April–30 September) is 1.40. Since our calculations show that the time-averaged probabilities in winter nights and in summer nights are approximately equal, therefore at Homestake the ratio between the numbers of Ar atoms produced in Cl by solar neutrinos during the winter and during the summer should be between 1 and 1.40 (1.40 if the MSW effect in the Sun transforms all solar ν_e s into ν_μ s, less than this otherwise).

After this paper was submitted for publication three further papers have appeared on the topic of resonant amplification of oscillations of atmospheric and solar neutrinos in Earth¹⁶⁻¹⁸.

This work was supported in part by the Adler Foundation for Space Research and the Technion Research Fund.

Received 11 August; accepted 8 December 1986.

1. Mikheyev, S. P. & Smirnov, A. Yu. *Nuovo cim.* **9C**, 17–26 (1986).
2. Wolfenstein, L. *Phys. Rev.* **D17**, 2369–2374 (1978); **D20**, 2634–2635 (1979).
3. Bahcall, J. N. *et al. Rev. mod. Phys.* **54**, 767–799 (1982).
4. Bahcall, J. N. *et al. Astrophys. J.* **292**, L79–L82 (1985).
5. Bethe, H. A. *Phys. Rev. Lett.* **56**, 1305 (1986).
6. Rosen, S. P. & Gelb, J. M. *Phys. Rev. D* **34**, 969–979 (1986).
7. Dar, A., Mann, A., & Melina Y. *Phys. Rev. D* (submitted).
8. Kuzmin, V. A. *JETP* **22**, 1051–1056 (1966).
9. Bahcall, J. N. *et al. Phys. Rev. Lett.* **40**, 1351–1354 (1978).
10. Hampel, W. in *Solar Neutrinos and Neutrino Astronomy (Homestake, 1984)* (eds Cherry, M. L., Fowler, W. A. & Lande, K.) 126–136 (AIP, New York, 1985).
11. Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Interiors* **25**, 297–356 (1981).
12. Dar, A. *Phys. Rev. Lett.* **51**, 227–230 (1983).
13. Gaisser, T. K. *et al. Phys. Rev. Lett.* **51**, 223–226 (1983).
14. Perkins, D. H. in *'Neutrino 84'* (eds Kleinknecht K. & Pashos, E. A.) 374–380 (World Scientific Publication Co., Singapore, 1984).
15. Davis, R. Jr *Proc. 7th Workshop Grand Unification*, Toyama (in the press).
16. Loscocco, J. M. *Phys. Rev. Lett.* **57**, 652–653 (1986).
17. Carlson, E. D. *Phys. Rev. D* **34**, 1454–1459 (1986).
18. Mikheyev, S. P. & Smirnov, A. Yu. in *1986 Massive Neutrinos* (eds Fackler, O. & Tran Thanh Van, J.) 355–372 (Kim Hup Lee Publishing Co., Singapore, 1986).

Icosahedral C_{60} : an aromatic molecule with a vanishingly small ring current magnetic susceptibility

V. Elser & R. C. Haddon

AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

Smalley and co-workers¹ have speculated that an icosahedral C_{60} molecule might have unusual magnetic properties. Here we report that the π -electron ring-current susceptibility of C_{60} is unusually small and sensitively dependent on the relative strengths of the two inequivalent bonds of the molecule. For equal bond strengths the susceptibility is only -0.21 that of benzene (with field normal to the plane of the six-membered ring). A 2% change in relative bond strengths is enough to change the sign of the susceptibility with the result that C_{60} is probably weakly diamagnetic. The shielding effect of the π -electrons is calculated to be less than 1 p.p.m. for an atom placed at the centre of C_{60} . A larger effect is expected for C_{60}^+ , which is calculated to be strongly diamagnetic.

Recently Smalley and co-workers¹⁻³ have produced evidence for the generation of C_{60} , which they formulated as a truncated icosahedron. Their suggestion that C_{60} might provide the first example of a spheroidal aromatic molecule (Fig. 1) has been supported by π -electron and total energy calculations⁴⁻¹⁰. A further suggestion by Smalley and co-workers¹ which (in certain circumstances^{11,12}) also bears on the aromatic character of C_{60} has yet to receive theoretical attention. They noted that "The inner and outer surfaces are covered with a sea of π electrons" and went on to suggest that "... the chemical shift in the NMR

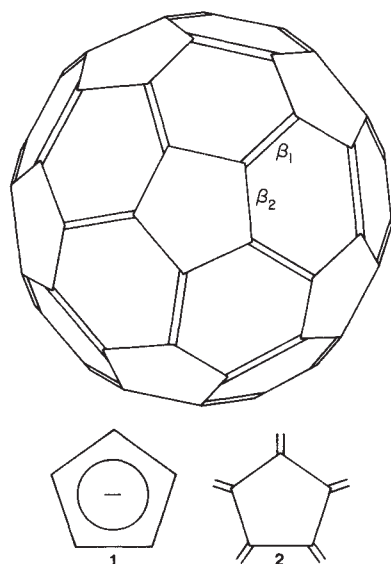


Fig. 1 Molecular structure of icosahedral C₆₀; cyclopentadienide anion (1); [5]radialene (2).

of the central atom [placed in the interior] should be remarkable because of the ring currents." This intriguing hypothesis is the subject of the present paper.

We have calculated the ring-current magnetic susceptibility of C₆₀ with the finite-field version of the original London theory¹³. Adopting a common bond length (with resonance integral (taken to be positive) equal to that of benzene (β)), we find the π -electron ring current magnetic susceptibility of C₆₀ to be -0.21 that of benzene ($\chi_{\pi}^{C_{60}}/\chi_{\pi}^{\text{benzene}}$ with field normal to the plane of the benzene six-membered ring (6-MR)). Thus the susceptibility of C₆₀ is much smaller than that of benzene and is paramagnetic (but see below). $\chi_{\pi}^{C_{60}}$ is isotropic so the rotationally averaged (solution) value is unchanged from the static value, whereas $\chi_{\pi}^{\text{benzene}}$ is decreased threefold. Scaling the resonance integral in C₆₀ for non-planarity (as in the 3-dimensional Hückel molecular orbital theory)^{10,14} would further reduce χ_{π} .

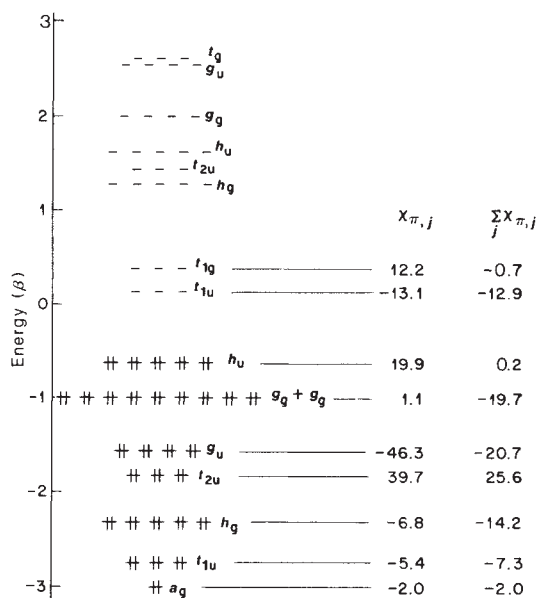


Fig. 2 Hückel molecular orbital²⁰, j , energy levels in units of β and π -orbital magnetic susceptibilities in units of $|\chi_{\pi}^{\text{benzene}}|$ calculated for C₆₀ with a common resonance integral. Occupation of an orbital is denoted by a solid vertical line; labels refer to irreducible representations of the icosahedral point group.

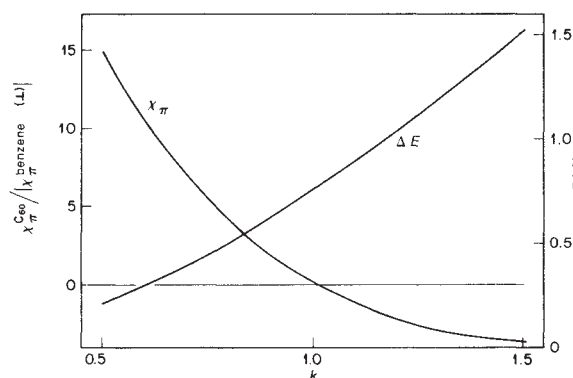


Fig. 3 π -Electron magnetic susceptibility as a function of the relative resonance integrals of C₆₀. Also shown is the behaviour of ΔE , the energy gap between the highest occupied and the lowest unoccupied molecular orbitals.

Figure 2 shows the contributions to the magnetic susceptibility by individual molecular orbitals ($\chi_{\pi,j}$). Many of the orbitals possess large susceptibilities, and the low molecular magnetic susceptibility is brought about by an essentially complete cancellation (see below). It has been predicted that C₆₀ will have an exceptionally high electron affinity¹⁰, and C₆₀⁻ is calculated to be strongly diamagnetic and C₆₀¹²⁻ weakly diamagnetic (Fig. 1). This contrasts with planar aromatic hydrocarbons and graphite, which show appreciable diamagnetism in the neutral state and a very strong ring-current paramagnetism when reduced¹⁵⁻¹⁹.

In a magnetically isotropic molecule, with π -electrons on the surface of a sphere, the ring-current contribution to the NMR chemical shift of a central atom takes the form $\delta_{RC} = 2\chi_{\pi}/r^3$, where r is the radius of the sphere and positive δ signifies deshielding. For C₆₀ with a radius of 2.4781 (where l is the carbon-carbon bond length, 1.4 Å), the relationship takes the form $\delta_{RC} = 0.0443 (\chi_{\pi}^{C_{60}}/|\chi_{\pi}^{\text{benzene}}|) (\beta/\text{kcal mol}^{-1})$ p.p.m.. Using the two empirical parameterizations for β (61.4(A) and 50.4(B) kcal mol⁻¹)¹⁵, we obtain for the chemical shifts (respectively), $\delta_{RC} = 0.57, 0.46$ p.p.m. (C₆₀); $-35.0, -28.8$ (C₆₀⁻); $-1.9, -1.5$ (C₆₀¹²⁻).

Although all the atoms in C₆₀ are related by symmetry, there are two types of bond: those at the junction of two 6-MRs (denoted by β_1) and those between a 5-MR and a 6-MR (β_2) (Fig. 1). If $\beta_1 = k\beta$ and $\beta_2 = \beta$, the ratio of the resonance integrals is given by $\beta_1/\beta_2 = k$. All the theoretical calculations find the 6/6-MR bonds to be shorter than the 5/6-MR bonds⁴⁻¹⁰. Thus $\beta_1 > \beta_2$ and, from the calculated bond lengths in the literature and an exponential dependence of the resonance integral on bond length²⁰, it is likely that $1.0 < k < 1.3$. Figure 3 shows χ_{π} for C₆₀ as a function of k : χ_{π} changes sign in the vicinity of $k = 1.02$. Given the estimates⁴⁻¹⁰ of the bond lengths in C₆₀ it is likely that k is larger than this and that the ring-current susceptibility in C₆₀ is weakly diamagnetic purely because of the relative bond strengths.

A useful reference point in understanding the completely unexpected magnetic properties of C₆₀ is provided by the cyclopentadienide anion (1) and [5]radialene (2)²¹. For the monocyclic 5-MR systems 1 and 2 the relative susceptibilities are given by $\chi_{\pi}^2/\chi_{\pi}^1 = 0.0015$. Thus it would seem that the magnetic response of C₆₀ is rather like that of [5]radialene. This point of view is complemented by the results in Fig. 3, where the paramagnetism increases as the energy gap (ΔE) decreases. The bond order in C₆₀ is large and thus a strong Larmor diamagnetism is expected (see below), but this is apparently cancelled by substantial Van Vleck paramagnetism of order $1/\Delta E$ ^{22,23}.

As a final comparison, we consider the classical Pauling²⁴ free-electron treatment of the ring-current magnetic susceptibility of C₆₀. If the molecule is treated as a spherical surface of 60 π -electrons and the susceptibility calculated from the Larmor

formula for free precession of the electrons in a magnetic field, we obtain $\chi_\pi = -60(e^2/4m_e c^2)\langle\rho^2\rangle$, where $\langle\rho^2\rangle$, the expectation value of the mean square radius of the electron orbits is given by $2r^2/3 = 4.09l^2$. Thus the free electron theory gives $\chi_{\pi C_{60}}/|\chi_{\pi \text{benzene}}| = -(60/6)(4.09) = -40.9$ (strong diamagnetism). The quantum-mechanical London calculation presented earlier shows that the correct value is less than one-hundredth of the classical value.

Although planar aromatic hydrocarbons and graphite have provided appropriate points of analogy for the π -electron energy structure of C_{60} , it is clear that the magnetic response of C_{60} is unlike that of any other molecule yet encountered.

Received 18 September; accepted 27 November 1986.

1. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162-163 (1985).
2. Zhang, Q. L. *et al. J. Phys. Chem.* **90**, 525-528 (1986).
3. Liu Y. *et al. Chem. Phys. Lett.* **126**, 215-217 (1986).
4. Bochavar, D. A. & Gal'pern, E. G. *Dokl. Akad. Nauk SSSR, Chem., Engl.* **209**, 239-241 (1972).
5. Davidson, R. A. *Theor. chim. Acta* **58**, 193-231 (1981).
6. Haymet, A. D. J. *Chem. Phys. Lett.* **122**, 421-424 (1985); *J. Am. chem. Soc.* **108**, 319-321 (1986).
7. Klein, D. J., Schmalz, T. G., Hite, G. E. & Seitz, W. A. *J. Am. chem. Soc.* **108**, 1301-1302 (1986).
8. Newton, M. D. & Stanton, R. E. *J. Am. chem. Soc.* **108**, 2469-2470 (1986).
9. Disch, R. L. & Schulman, J. M. *Chem. Phys. Lett.* **125**, 465-466 (1986).
10. Haddon, R. C., Brus, L. E. & Raghavachari, K. *Chem. Phys. Lett.* **125**, 459-464 (1986).
11. Haddon, R. C. *Aust. J. Chem.* **30**, 1-22 (1977).
12. Haddon, R. C. *J. Am. chem. Soc.* **101**, 1722-1728 (1979).
13. London, F. J. *Phys. Rad.* **8**, 397-409 (1937).
14. Haddon, R. C. *J. Am. chem. Soc.* (submitted).
15. Haddon, R. C. *Tetrahedron* **28**, 3613-3634, 3635-3655 (1972).
16. Kuwajima, S. & Soos, Z. G. *J. Am. chem. Soc.* **108**, 1707-1708 (1986).
17. DiSalvo, F. J., Safran, S. A., Haddon, R. C., Waszczak, J. V., Fischer, J. E. *Phys. Rev. B* **20**, 4883-4888 (1979).
18. Haigh, C. W. & Mallion, R. B. in *Progress in NMR Spectroscopy* (eds Emsley, J. W., Feeney, J. & Sutcliffe, L. H.) Vol. 13, 303-344 (Pergamon, Oxford, 1979).
19. Garratt, P. J. *Aromaticity* (Wiley, New York, 1986).
20. Salem, L. *Molecular Orbital Theory of Conjugated Systems*, Ch. 3 (Benjamin, New York, 1966).
21. Aihara, J.-I. *Bull. chem. Soc. Japan* **53**, 1751-1752 (1980).
22. McWeeny, R. *Molec. Phys.* **1**, 311-321 (1958).
23. Elser, V. & Haddon, R. C. *Phys. Rev. A* (submitted).
24. Pauling, L. *J. chem. Phys.* **4**, 673-677 (1936).

A new method for measuring the surface energy of solids

K. Kendall, N. McN. Alford & J. D. Birchall

ICI New Science Group, PO Box 11, The Heath, Runcorn, Cheshire WA7 4QE, UK

Although the surface tension of liquids has been understood and measured since the time of Young¹ and Laplace², the surface energy of solids has eluded understanding and evaded measurement, despite its postulated importance to catalysis, crystal growth, colloidal behaviour, sintering and fracture. It is salutary to reflect that solid surface energies are even now more readily determined by theoretical argument than by experiment. Here we show that solid surface energies can be evaluated experimentally by measuring the elastic modulus of submicrometre powder assemblies, knowing the particle diameter, elastic modulus and volume fraction.

The problem is that solids should display higher surface energies than liquids, but their atoms are relatively immobile, so that stresses in the surface may exist in addition to the surface tension of the solid. Attempts to distinguish the genuine solid surface tension from such spurious stresses have not been satisfactory^{3,4}. For the same reason experiments to find the surface energy of the solid by measuring the enhanced vapour pressure⁵, increased solubility⁶, enlarged heat of dissolution⁷ or lowered melting point⁸ of powders have had only limited success; other energy terms shroud the surface contribution to such an extent that the very existence of solid surface energy has been doubted⁹.

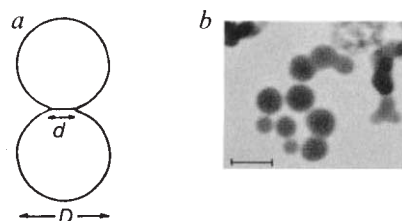


Fig. 1 a, Two elastic, smooth spheres in contact under zero external load form a contact spot of diameter d . b, Transmission electron micrograph of silica particles (OX50, Degussa). Scale bar, 0.1 μm .

Similar obstacles have dogged the cleavage method, the most popular way of estimating solid surface energies since Griffith⁹ proposed his criterion of brittle fracture and Obreimoff¹⁰ showed the importance of gaseous contamination. Although cleavage experiments have given results close to the theoretically expected values¹¹, it is more usual to obtain inordinately high figures because plastic flow occurs under the high stresses near the crack tip, even for brittle materials, causing a large and indeterminate energy demand¹².

Our new method avoids this difficulty because it is an equilibrium measurement which does not involve cleavage. Instead it is a development of the method devised for rubber, measuring the equilibrium diameter of the contact between two solid bodies held together by surface forces¹³. The interfacial energy between two plane surfaces of unit area was defined as Γ , the energy released on bringing the surfaces from infinity into contact. For identical solids, Γ was twice the surface energy of each piece, $\Gamma = 2\gamma$.

Experimentally, it proved better to use spheres rather than plane surfaces and to measure the enlargement of the Hertz¹⁴ elastic contact circle by the attractive surface forces¹⁵⁻¹⁹. The interface energy Γ was then calculated from the equation

$$d^3 = 3(1-\nu^2) \frac{D}{E} \left[W + \frac{3}{4}\pi D\Gamma + \sqrt{\frac{3}{2}\pi DWT + \left(\frac{3}{4}\pi D\Gamma\right)^2} \right] \quad (1)$$

where d was the contact spot diameter, W the external load on the spheres, D the sphere diameter, E its Young's modulus and ν its Poisson's ratio. It was demonstrated that this method gave a better approach to equilibrium than a fracture test on the same geometry¹³. For 44 mm diameter rubber spheres the interfacial energy was measured to be $\Gamma = 71 \text{ mJm}^{-2}$ falling to 6.8 in water and to less than 1 mJm^{-2} in a 0.01 molar solution of sodium dodecyl sulphate. These values, inserted in Young's equation, were consistent with contact angle determinations. The main problem was surface roughness, inhibiting molecular contact and restricting the method to 'smooth' solids such as rubber or mica²⁰ which satisfied a smoothness criterion²¹ depending largely on the centre line average roughness and the elastic modulus of the sample. Large spheres of ceramics would require almost atomically smooth surfaces to produce an adhesive contact.

In this report we recognize that this restriction of solid smoothness applies only to large bodies. It has previously not been observed that equation (1) should apply to all solids if the sphere size is made sufficiently small. In the limit, a single atom cannot be rough. Similarly, a 10 nm diameter particle should appear smooth as far as adhesion is concerned. Although the upper bound to this argument has not been defined (because Fuller and Tabor²¹ only addressed the contact of plane surfaces, following Johnson²²) it is reasonable to suppose that equation (1) should hold for ceramic grains less than 1 μm in size. Particles of this kind are found to adhere to a considerable degree²³. The snag in using such particles to measure surface energy is that the contact spot is only tens of nm in diameter. Electron microscopy proved useful in defining the contact for metal and polymer latex spheres^{24,25}. Here it is shown that the contact spot

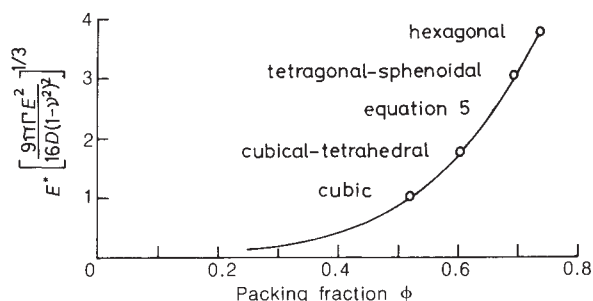


Fig. 2 Calculated elastic modulus for four regular sphere packings for comparison with equation (5).

size may be determined from the elastic modulus of a powder assembly²⁶ allowing the interfacial energy Γ to be calculated.

Consider two equal spheres in contact under zero external load (Fig. 1a). The diameter of the contact spot from equation (1) is

$$d = \left[\frac{9\pi}{2E} \Gamma D^2 (1 - \nu^2) \right]^{1/3} \quad (2)$$

This can be measured by applying a small test load P and following the elastic deformation δ of the spheres since¹³

$$P/\delta = Ed/2(1 - \nu^2) = \left[\frac{9\pi\Gamma E^2 D^2}{16(1 - \nu^2)^2} \right]^{1/3} \quad (3)$$

Although the deformation at a single contact is minute, it can be magnified by stacking the spheres end to end in a regular cubic packing, whose elastic modulus E^* from equation (3) is

$$E^* = P/d\delta = \left[\frac{9\pi\Gamma E^2}{16D(1 - \nu^2)^2} \right]^{1/3} \quad (4)$$

To judge the effect of changing the packing from simple cubic, the elastic modulus was also calculated for hexagonal close packing and for two other regular packings of uniform spheres. It was found that the elastic modulus increased with packing fraction ϕ (volume of spheres/volume of assembly) along a single curve (Fig. 2) described by

$$E^* = 17.1 \phi^4 \left[\frac{\Gamma E^2}{D} \right]^{1/3} \quad (5)$$

Qualitatively, this ϕ^4 dependence of E may be understood in terms of two effects; first the ϕ^2 dependence of modulus on the density of particle packing²⁷; second, a further ϕ^2 dependence of modulus on the coordination number of each sphere²⁸. Based on this simple reasoning, it was proposed that irregular sphere packings should also fall on this same curve.

The following experiments were designed to test this hypothesis and then to use equation (5) to measure the interface energy Γ of ceramic particles. Silica (OX50, Degussa, $D = 0.05 \mu\text{m}$, $E = 70 \text{ GPa}$) (Fig. 1b) and titania (RCR2, Tioxide, $D = 0.23 \mu\text{m}$, $E = 280 \text{ GPa}$) were used to make the particle assemblies because they were nearly spherical and of relatively narrow size distribution. The particles were mixed in a polymer solution (30% polyvinyl alcohol (KH 17 s, Nippon Gohsei) in water) pressed to a sheet, dried, cut into strips, then heated to 500°C to remove the polymer. The resulting particle assemblies were bend tested to determine Young's modulus E^* which was plotted against packing fraction ϕ derived from the weight and volume of the strip.

The measurements for titania (Fig. 3a) gave a good fit to the ϕ^4 relation of equation (5) and correspond to a value of $\Gamma = 0.6 \text{ Jm}^{-2}$ reasonably close to the theoretical estimate of $\Gamma = 0.8 \text{ Jm}^{-2}$ for the 110 plane of rutile at 20°C ²⁹. A powder compact made by pressing in a steel die gave a result near the curve, indicating that the modulus was not strongly affected by the preparation method. Possible errors in the experimental results

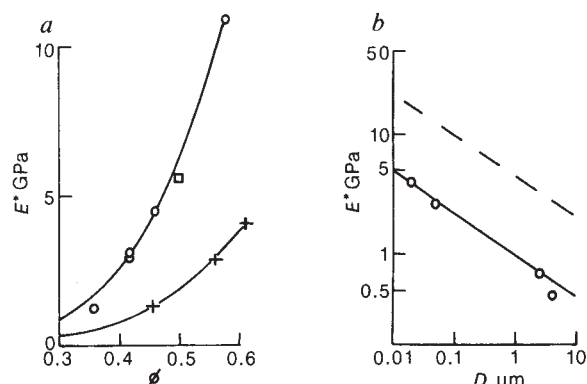


Fig. 3 a, Elastic modulus results for particle assemblies compared with equation (5). $\circ$, Titania powder, mixed in polymer solution. $\square$, Titania powder, die pressed, $\Gamma = 0.6 \text{ Jm}^{-2}$. $+$, Silica powder, $\Gamma = 0.05 \text{ Jm}^{-2}$. b, Results for silica powders of different particle diameter compared to equation (5). The broken line shows equation (5) plotted for $\Gamma = 1.16 \text{ Jm}^{-2}$, a theoretical estimate for pure silica²⁵. The other line shows equation (5) plotted for $\Gamma = 0.05 \text{ Jm}^{-2}$.

stem from inaccuracy of particle diameter determination, the spread of particle sizes in the powder and anisotropy of particle shape and elasticity.

Results for silica also fitted equation (5) but corresponded to $\Gamma = 0.05 \text{ Jm}^{-2}$, much lower than the theoretical 1.16 Jm^{-2} for the 110 plane of β cristobalite²⁹. This low value was confirmed by additional measurements on a series of pure silica powders of different diameter, results which supported the $D^{-1/3}$ dependence of modulus (Fig. 3b). The low result for Γ was unchanged when the silica assembly was immersed in water, suggesting that the silica surface was already saturated by atmospheric contamination. Heating the silica to 700°C gave a weight loss of almost 1% and increased the modulus by a factor of nearly 2, indicating that water was being driven off to increase the surface energy of the silica. On cooling and rewetting, the modulus fell back again. This observation is important to ceramics, which require low surface energy to ease powder compaction³⁰ but a high surface energy to drive sintering at elevated temperatures.

In conclusion, the surface energy of solids may be measured by determining the elastic modulus of fine powder assemblies, knowing the packing fraction, the diameter and the elastic properties of the grains. Although there are difficulties because many powders are not spherical, uniform or isotropic, it is believed that this new method may make the surface energy of solids as experimentally accessible as that of liquids, adding to our knowledge and application of surface phenomena.

Received 11 August; accepted 16 December 1986.

- Young, T. *Phil. Trans. R. Soc.* **95**, 65–87 (1805).
- Laplace, P. S. *Mécanique Céleste* Suppl. Book 10 (translated by Bowditch, N.) 685–1018 (Chelsea Publishing, New York, 1966).
- Chapman, J. C. & Porter, H. L. *Proc. R. Soc. A* **83**, 65–68 (1910).
- Berggren, B. *Ann. Phys.* **44**, 61–80 (1914).
- Bikerman, J. J. *Phys. Status Solidi* **10**, 3–25 (1965).
- Alexander, G. B. *J. phys. Chem.* **A171**, 1563–1564 (1957).
- Benson, G. C., Schreiber, H. P. & van Zeggeren, F. *Can. J. Chem.* **34**, 1553–1556 (1956).
- Sill, R. C. & Skapski, A. S. *J. chem. Phys.* **24**, 644–651 (1956).
- Griffith, A. A. *Phil. Trans. R. Soc. A* **221**, 163–198 (1920).
- Obreimoff, J. W. *Proc. R. Soc. A* **127**, 290–297 (1930).
- Gilman, J. J. *J. appl. Phys.* **31**, 2208–2218 (1960).
- Lawn, B. R. & Wilshaw, T. R. *Fracture of Brittle Solids*, Ch. 6 (Cambridge University Press, Cambridge, 1975).
- Johnson, K. L., Kendall, K. & Roberts, A. D. *Proc. R. Soc. A* **324**, 301–313 (1971).
- Hertz, H. J. *Reine angew. Math.* **92**, 156–171 (1882).
- Tabor, D. *J. Colloid. Interface Sci.* **58**, 2–13 (1977); **67**, 380 (1978); **73**, 294 (1980).
- Derjaguin, B. V., Muller, V. M. & Toporov, Y. P. *J. Colloid Interface Sci.* **53**, 314–326 (1975); **67**, 378–379 (1978); **73**, 293 (1980).
- Muller, V. M., Derjaguin, B. V. & Toporov, Y. P. *Colloids Surfaces* **7**, 251–259 (1983).
- Pashley, M. D. *Colloids Surfaces* **12**, 69–77 (1984).
- Horn, R. G., Israelachvili, J. N. & Pribac, F. *J. Colloid Interface Sci.* (in the press).
- Maugis, D. & Barquins, M. in *Adhesion and Adsorption of Polymers* (ed. Lee, L.) 203–277 (Plenum, New York, 1980).
- Fuller, K. N. G. & Tabor, D. *Proc. R. Soc. A* **345**, 327–342 (1975).

22. Johnson, K. L. *Proc. IUTAM Symp. on The Mechanics of Contact between Deformable Bodies* (eds de Pater & Kalker) 26-40 (Delft University Press, 1975).
23. Kendall, K., Alford, N. McN. & Birchall, J. D. *Spec. Ceram.* 8, 255-265 (Institute of Ceramics, Stoke-on-Trent, 1986).
24. Easterling, K. E. & Thölen, A. R. *Acta met.* 20, 1001-1008 (1972).
25. Kendall, K. & Padgett, J. C. *Int. J. Adhesion Adhesives* July, 149-154 (1982).
26. Kendall, K., Alford, N. McN. & Birchall, J. D. *Proc. R. Soc.* (submitted).
27. Holliday, L. *Composite Materials* Ch. 2 (Elsevier, London, 1966).
28. Norman, L. D. & Maust, R. *Bull. U.S. Bur. Mines* 58, 35-48 (1971).
29. Bruce, R. H. *Sci. Ceram.* 2, 359-367 (1966).
30. Kendall, K. *Nature* 319, 203-205 (1986).

Mechanical instability of gels at the phase transition

Toyoichi Tanaka, Shao-Tang Sun, Yoshitsugu Hirokawa, Seiji Katayama, John Kucera, Yoshiharu Hirose & Takayuki Amiya

Department of Physics and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Polymer gels, consisting of a cross-linked polymer network immersed in liquid, undergo a volume phase transition: when external conditions such as temperature or solvent composition change, a gel reversibly swells or shrinks, but does so discontinuously¹⁻⁵. The volume change at the transition can be as large as a factor of one thousand², and the phenomenon occurs in all gels^{6,7}. The equilibrium aspects of the phase transition have been extensively studied, but its kinetics have not yet been fully explored. In particular, the appearance of patterns on the originally smooth surface of a gel during the transition makes the kinetic process difficult to understand. Here we elucidate the physical basis underlying the formation and evolution of the pattern.

When a polymer gel undergoes an extensive swelling, a beautiful, regular pattern appears on the surface of the gel (as shown on this week's cover). At the beginning, the pattern is extremely fine, having a texture similar to that of a frosted glass. As time goes on, the units of pattern coalesce, doubling their characteristic size. When the unit size becomes comparable to the size of the gel, the pattern gradually disappears. As the gel approaches equilibrium, it regains the shape identical to the original one. Finally the gel becomes totally homogeneous, and the process comes to an end.

The pattern is formed in all the gels we have studied, that is, some polyvinyl gels and gels made of natural polymers. We

chose ionized acrylamide gels because of their relative ease of handling. Copolymeric gels of acrylamide (700 mM) and sodium acrylate (0-32 mM) are prepared by a standard method using free radical polymerization initiated by ammonium persulphate and catalysed by tetramethylethylenediamine¹. The osmotic pressure of counterions from sodium acrylate exerts an internal osmotic pressure and causes the gel to expand when placed in water. The volume expansion increases with the amount of sodium acrylate. The gel has a radius ~2 cm before swelling and 8 cm after complete swelling. The gel has a perfectly smooth surface before swelling, and as the gel is placed in water an extremely fine pattern appears on the surface⁸. As time goes on, the unit size of the pattern becomes larger by coalescence of units into a larger unit, which is accomplished by conversion of side lines into thorns followed by the disappearance of thorns. Eventually the size of the unit becomes comparable to that of the gel radius, when the pattern disappears completely. It continues to swell to the final volume until it regains the spherical shape and totally homogeneous density.

Careful observation shows that the pattern consists of numerous line segments of cusps into the gel. These cusps are not due to breaking of the gel, nor are the representative of a shrunken portion of the gel. The line is a result of shear bending of a homogeneously swollen gel surface. Indeed, when the surface is sliced off in a thin layer, the pattern disappears instantly, and the slice becomes homogeneous. If the cusps were due to local shrinking of the gel, they would take a long time to disappear through the collective diffusion process. At each swelling stage, the pattern appears to be a hexagonal lattice, but it is quite irregular. A similar pattern is observed on the surface of the acrylamide gel formed in a Petri dish and swollen in water (Fig. 1). The patterns are formed only when swelling is extensive: if the swelling is small, the surface of the gel remains smooth throughout the entire process.

The pattern formation may be qualitatively explained in the following way. The kinetic process of the swelling of a gel is governed by the collective diffusion of the polymer network into a solvent⁹. At first, a very thin surface layer is swollen. This layer is under a mechanical constraint, namely, the outer surface of the layer is free to expand, whereas the inner surface is fixed to the core of the gel. Thus, the layer is under opposing demands on the upper and lower surfaces, one to expand and the other to remain unswollen. When the osmotic pressure is small, these opposing forces are resolved by stretching the gel unidirectionally perpendicular to the surface. When the osmotic pressure is large the outer surface is forced to buckle. The characteristic wavelength of the pattern must be proportional to the thickness

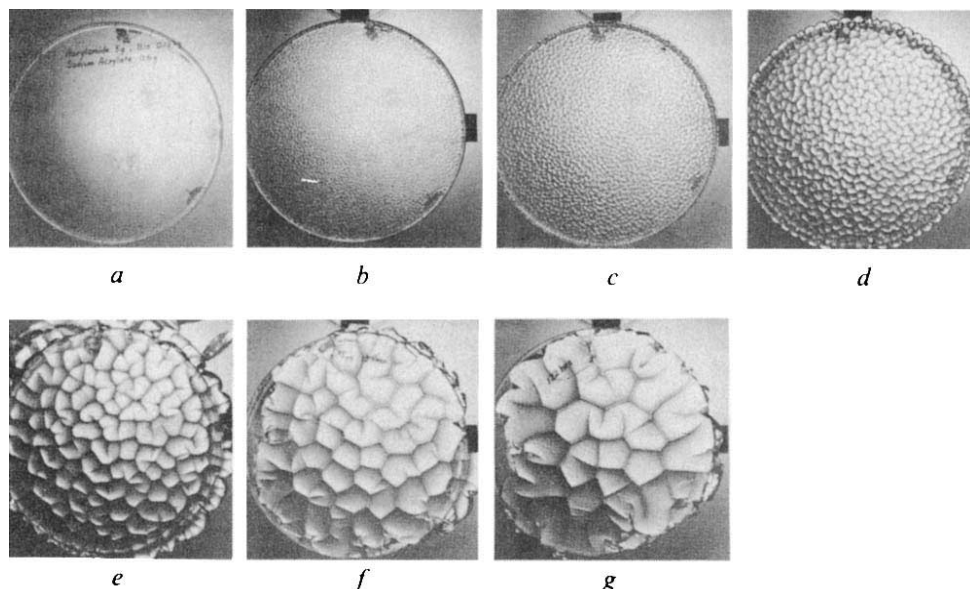


Fig. 1 An ionized acrylamide gel formed in a Petri-dish is allowed to swell in water. An extremely fine pattern appears on the free surface of the disk gel, and evolves with time ($a \rightarrow g$).

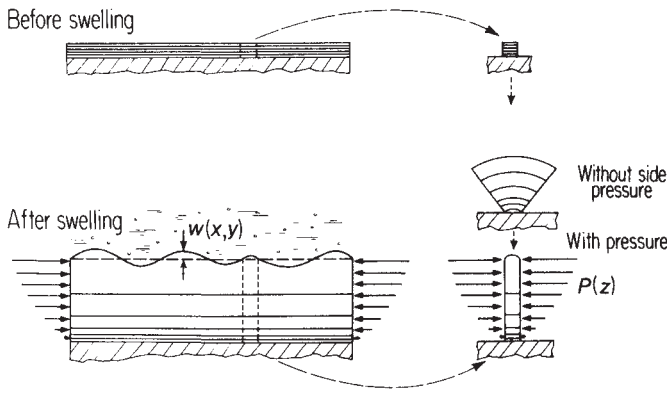


Fig. 2 The model of a gel on which the elastic deformation energy is calculated in the text. The gel slab is treated as a combination of thin layers. The lowest layer is mechanically fixed to a film, but the top layer is free to expand except for its boundary, which is fixed only in the horizontal direction.

of the swollen layer, as that is the only relevant length scale. As time goes on, the thickness of the swollen layer increases, as does the wavelength of the pattern.

To test the hypothesis presented above, we have carried out several demonstrative experiments. Gel slabs of thickness from 0.2–5 mm are prepared with one surface free, and the other surface covalently cross-linked to a film whose surface has polymerization-active chemical groups. When the gel slab is immersed in water, a pattern is formed and evolves in a similar way to the swelling of the spherical gel as shown before. But the evolution stops because the mechanical constraint is permanent. The final wavelength is proportional to the thickness of the gel slab.

We have further investigated the effect of internal osmotic pressure of the gel on the pattern formation. The relative amount of sodium acrylate is used to vary the osmotic pressure of counterions. There is a critical osmotic pressure above which distinct patterns appear, and below which the gel surface remains smooth, and this pressure is independent of the gel thickness.

We now theoretically analyse the phenomena using a logic similar to that introduced by Euler to describe the mechanical instability of a rod under a compressional force¹⁰. Consider a square slab of gel, with its upper surface free, but its lower surface mechanically fixed (see Fig. 2). When the gel is swollen the local degree of swelling monotonically increases from the fixed surface to the other free surface. The compressional pressure is a maximum at the top and decreases monotonically towards the lower surface. Let us assume that there appears a sinusoidal fluctuation in the gel, represented by a vertical displacement $\Delta z = w(x, y, z)$, where we define the coordinate system so that the z -axis is perpendicular to the gel surface. Assume further that the displacement changes monotonically with z . Then

$$w(x, y, z) = w(x, y) \sin\left(\frac{\pi z}{2h}\right) \quad (1)$$

The mechanical potential energy of the gel associated with this shear deformation is given by

$$V = \iiint \left\{ \frac{1}{2} \left[\frac{Ez^2}{9} (\Delta w)^2 - P(\nabla w)^2 + \frac{E}{h^2} w^2 \right] \right\} dr \quad (2)$$

The first term represents the bending energy of the gel, the second term the energy due to compression, and the last term represents the energy associated with the local vertical stretching. By Fourier-transforming the displacement

$$w(x, y) = \sum A_{\mathbf{k}} e^{i\mathbf{k}\mathbf{r}} \quad (3)$$

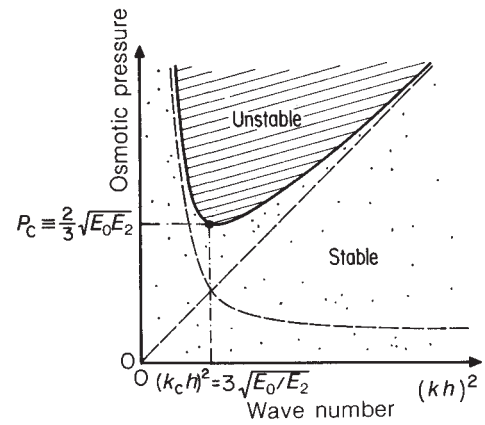


Fig. 3 The condition for instability is plotted for two independent parameters. The vertical axis represents the osmotic pressure and the horizontal axis denotes the wave-number of the fluctuations. The boundary for the instability depends on Young's modulus and the gel thickness.

the mechanical potential energy becomes a sum of independent modes.

$$V = \frac{L^2 h}{2} \sum \left\{ \frac{E_2}{9} (kh)^4 - P(kh)^2 + E_0 \right\} A_{\mathbf{k}} A_{-\mathbf{k}} \quad (4)$$

where

$$E_i = \int E(z) \left(\frac{z}{h}\right)^i \sin^2\left(\frac{\pi z}{2h}\right) \frac{dz}{h} \quad (i=0, 2) \quad (5)$$

and

$$P = \int P(z) \sin^2\left(\frac{\pi z}{2h}\right) \frac{dz}{h} \quad (6)$$

Modes having a negative value for the term in brackets in equation (4) become unstable, because then the increase in amplitude $A_{\mathbf{k}} A_{-\mathbf{k}}$ leads to lower energy. The instability condition

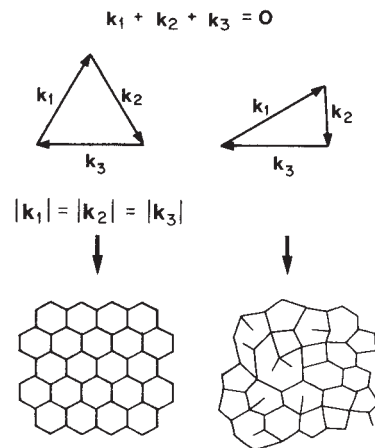


Fig. 4 A wave consisting of three standing plane waves having wave vectors $\mathbf{k}_1$, $\mathbf{k}_2$ and $\mathbf{k}_3$ of the same magnitude, which satisfy $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$ constructs a hexagonal pattern. This situation is similar to the instability problem of the hexagonal Benard cells observed in a fluid layer heated from the bottom¹¹. In the actual observation, however, the pattern is found to be not exactly hexagonal. Because the osmotic pressure is higher than the critical value P_c , more than one mode become unstable and grow at the same time. A combination of three modes having different wave numbers can now construct a closed, deformed triangular relationship. In such cases the pattern may be distorted from the perfect hexagonal pattern.

is given by

$$P > E_2 \frac{(kh)^2}{9} + E_0 \frac{1}{(kh)^2} \\ = \left(E_2^{1/2} \frac{kh}{3} - E_0^{1/2} \frac{1}{kh} \right)^2 + \frac{2}{3} (E_2 E_0)^{1/2} \quad (7)$$

The critical pressure for instability is different for each mode, and the first mode that becomes unstable when the osmotic pressure increases is given by

$$k_c = \frac{2\pi}{\lambda_c} = \left(\frac{E_0}{E_2} \right)^{1/4} \frac{\sqrt{3}}{h} \quad (8)$$

and the critical osmotic pressure is

$$P_c = \frac{2}{3} (E_2 E_0)^{1/2} \quad (9)$$

The predictions that the wavelength of the pattern is proportional to the thickness of the gel, and that the critical pressure is independent of the gel thickness agree with the experimental results (see Fig. 3).

Once a mode becomes unstable and its amplitude increases, it is necessary to take into account the higher order terms.

$$V = \frac{K(k)}{2} \sum |A_k|^2 - \frac{c}{2} \sum A_{k_1} A_{k_2} A_{k_3} + \frac{d}{4} \sum |A_{k_1}|^2 |A_{k_2}|^2 \quad (10)$$

where the second summation is taken for $k_1 + k_2 + k_3 = 0$. This potential energy is minimized for a combination of three waves having wavevectors constituting an equilateral triangle relationship. Such a wave represents a hexagonal pattern (Fig. 4).

This work was supported by NSF, grant DMR 83-10522.

Received 2 September 1986; accepted 2 January 1987.

1. Tanaka, T. *Phys. Rev. Lett.* **40**, 820-823 (1978).
2. Tanaka, T. *et al. Phys. Rev. Lett.* **45**, 1636-1639 (1980).
3. Ilavsky, M. *Macromolecules* **15**, 782-788 (1982).
4. Tanaka, T., Nishio, I., Sun, S.-T. & Ueno-Nishio, S. *Science* **218**, 467-469 (1982).
5. Katayama, S. & Ohata, A. *Macromolecules* **18**, 2781-2782 (1985).
6. Hirokawa, Y., Tanaka, T. & Sato, E. *Macromolecules* **18**, 2782-2783 (1985).
7. Amiya, T. & Tanaka, T. *Macromolecules* (in the press).
8. Hirokawa, Y. *et al. in Physical Optics of Dynamic Phenomena and Processes in Macromolecular Systems* (ed. Sedláček, B.) 197-204 (de Gruyter, Berlin, 1985).
9. Tanaka, T. & Fillmore, D. J. *J. chem. Phys.* **70**, 1214-1218 (1979).
10. Landau, L. D. & Lifschitz, E. M. *Teoriya Uprugosti* (Nauka, Moscow, 1967).
11. Haken, H. *Synergetics, An Introduction* (Springer, Berlin, 1983).

Anomalous $^{234}\text{U}/^{238}\text{U}$ ratios in deep-sea hydrothermal deposits

Jean Louis Reyss, Nathalie Lemaitre, Philippe Bonté & Didier Franck

Centre des Faibles Radioactivités, Laboratoire mixte CNRS-CEA, Avenue de la Terrasse, 91190 Gif sur Yvette, France

Although hydrothermal ferromanganese deposits often display variable $^{234}\text{U}/^{238}\text{U}$ ratios¹⁻⁵, published values generally do not differ by more than 20% from the seawater value of their hydrogenous counterparts⁶. Here we report very high $^{234}\text{U}/^{238}\text{U}$ ratios in hydrothermal manganese crusts recovered from the Sanghihe island arc system in the West Pacific. Such ratios, about twice that of normal seawater, confirm a low-temperature hydrothermal supply of uranium, which is best explained by leaching of the underlying rocks by the hydrothermal fluids and deposition together with the manganese oxides before mixing with seawater. This suggests that the $^{234}\text{U}/^{238}\text{U}$ in seafloor deposits may be used as a tracer of low-temperature hydrothermal reactions in the ocean crust.

During the past 10 years, major hydrothermal deposits have been discovered in association with both mid-ocean ridge systems and also island arc systems such as the South-West Pacific island arc (SWPIA)⁷. In the search for new hydrothermal sites

associated with arc volcanoes, we attempted numerous dredges during the cruise 'ESTASE-2' of the RV *J. Charcot* on the flanks of several seamounts in the two arc systems of Sanghihe and Talau ridges adjacent to the Cotabato and Philippine Trenches, South Mindanao Island. U-series isotopes were shown to provide good tracers for the rock-weathering processes as well as dating tools for submarine deposits⁸. Here we report chemical and radiochemical data obtained on two dredge hauls from the western flank of a volcano that rises 2,300 m above a 5,000 m sea floor in the northern part of the Sanghihe Ridge (5°02' N, 125°15' E). Dredge DR4 (3,200-m depth) collected several large fragments of manganese crust 10-15 cm thick and DR11 (2,800 m depth) obtained similar Mn crusts, together with rocks coated by 1-2 mm of Fe-Mn oxides, pieces of manganese breccia and sediments.

The manganese deposits consist of alternating layers of compact and granular material similar to the crusts described by Moore and Vogt⁴ in the Galapagos Spreading Center (GSC) and by Moorby *et al.*⁹ in the SWPIA. Several samples contain lenticular areas of yellow to green or red smectite. Mineralogically, all the samples, which were not allowed to dry, contain a quite-well-crystallized todorokite in contrast to the hydrogenous incrustations that generally consist of δMnO_2 . This todorokite is very unstable and heating results in a more or less complete conversion of the 10-Å to the 7-Å peak attributable to birnessite (Mellin *et al.*, in preparation). X-ray diffraction and chemical data on the smectite associated with the Mn oxides are very similar to those of the hydrothermal nontronite in the Galapagos Mounds¹⁰.

Table 1 Chemical composition

Sample depth	Na (%)	Mg (%)	Al (%)	Si (%)	K (%)	Ca (%)	Mn (%)	Fe (%)	Ni (p.p.m.)	Co (p.p.m.)
DR4A/3 25-55 mm	3	1.1	<	0.5	0.9	1.5	47	0.022	30	7.7
DR4B/3										
11-23 mm	3.2	1.9	0.4	0.8	0.9	1.5	51	0.17	27	7.6
45-50 mm	3.1	1.7	0.5	0.9	0.9	1.3	51	0.02	21	7.5
DR4C/F 35-45 mm	1.7	1.6	0.3	1.3	1.1	1.3	50	0.78	20	12.5
DR11/E 10-20 mm	2.3	1.7	0.2	6.4	1.2	1.1	41.3	4.7	15	13
DR11/B 0.00-0.06 mm	2.1	1.3	2.0	7.3	0.5	1.8	12.9	18.5	4560	950
Sediment DR11	2.7	2	8.4	21	1.5	2.3	0.6	5.8	165	22
Nontronite DR11/A	1.2	2.1	0.7	21.4	2.1	0.5	2.3	19.1	<	6.5
Galapagos Mounds ⁵	2.3			1.1			43.6		61	6.4
SWPIA hydroth. Mn crust ⁸	2.9	1.1	0.08		0.55	1.3	47	0.02	14	7
Hydrogenous Mn endmember ¹¹	1.04	1.04	1.18	5.22	0.49	2.6	22.2	19.0	5500	1300
Suboxic diagenetic Mn end member ¹¹	3.28	1.38	0.75	1.63	0.62	1.25	48	0.49	4400	35

Table 2 Radiochemical data

Sample depth	U (p.p.m.)	Th (p.p.m.)	U/Th	$^{234}\text{U}/^{238}\text{U}$	^{230}Th (d.p.m. g ⁻¹)	^{227}Th as ^{231}Pa (d.p.m. g ⁻¹)	Age $^{230}\text{Th}/^{234}\text{U}$ Ky	Age $^{231}\text{Pa}/^{235}\text{U}$ Ky
DR4A/3 25–55 mm	4.29 ± 0.16	<0.02	>200	1.50 ± 0.04	2.39 ± 0.10	0.102 ± 0.016	72 ± 6	59
DR4B/3 11–23 mm	3.60 ± 0.07	0.12 ± 0.04	30	1.45 ± 0.08	2.19 ± 0.08	0.096 ± 0.020	86 ± 6	75
45–50 mm	6.57 ± 0.14	0.02 ± 0.01	>200	1.44 ± 0.02	3.81 ± 0.11	0.176 ± 0.019	81 ± 4	76
DR4C/F 25–40 mm	3.47 ± 0.17	0.07 ± 0.02	49	1.87 ± 0.03	2.089 ± 0.005	0.076 ± 0.006	59 ± 2	51
DR11/E 10–20 mm	2.46 ± 0.06	0.012 ± 0.004	205	2.25 ± 0.03	0.530 ± 0.010	0.023 ± 0.004	14.8 ^{+0.8} –0.9	15.9
DR11/B 0.00–0.06 mm	11.3 ± 0.5	35.1 ± 2.4	0.32	1.18 ± 0.04	270 ± 10	91 ± 4	excess	excess
Sediment DR11	0.95 ± 0.05	2.3 ± 0.2	0.41	1.06 ± 0.09	3.22 ± 0.20	0.450 ± 0.050	excess	excess
Nontronite DR11/A	0.20 ± 0.04	0.17 ± 0.03	1.17	1.18 ± 0.09	0.238 ± 0.019	n.m.	excess	

n.m., Non-measurable.

The thin Fe–Mn coating of sample DR11/B consists of δMnO_2 . This coating covers a piece of fresh dacite-andesite that was probably broken from the basement during dredging. The sediment itself, present in large amount in the dredge haul DR11 and adhering to the Mn crusts, consists mainly of detrital components: feldspar, quartz, kaolinite and montmorillonite.

Results of chemical analysis on the samples are given in Table 1; Na, Mg, Al, Si, K, Ca, Mn were obtained with an EDAX/SEM system, and Fe, Ni and Co were measured by INAA. The chemical composition of hydrothermal, hydrogenous and diagenetic manganese-rich deposits of deep-sea origin are also shown for comparison. The thin coating of DR11/B has the composition of typical hydrogenous deposits¹¹ with a Mn/Fe ratio of ~1 and a relatively high Co and Ni content. On the other hand, the chemical compositions of samples DR4 and DR11/E are similar to the hydrothermal crusts of GSC and SWPIA with high Mn (about 50%), low Fe and very low trace element (for example Ni and Co) content. These characteristics contrast with DR11/B and also with the diagenetic manganese deposit. Sample DR11/E resembles DR4 but has rather more Si and Fe, in the same ratio as in the nontronite; its composition may result from the mixing of the latter with the Mn-phase. The nontronite sample has very similar composition to that reported by Corliss *et al.*¹⁰ in the Galapagos Mounds.

We have measured U and Th series isotopes in the four different phases present in the dredged samples: sediment, nontronite, manganese crust and the Fe–Mn coating. U and Th were measured by alpha-spectrometry by the same procedure as described in Reyss *et al.*¹² For all the samples, the measurements were repeated without the ^{232}U spike to ensure the $^{234}\text{U}/^{238}\text{U}$ and the $^{227}\text{Th}/^{230}\text{Th}$ ratios were accurately established. For sample DR11/B, ^{210}Pb and ^{226}Ra were measured by gamma-spectrometry with a well-type Ge detector. Radiochemical data are presented in Table 2. The ages, deduced from ^{230}Th ingrowth towards radioactive equilibrium with ^{234}U and from ^{227}Th as ^{231}Pa ingrowth towards ^{235}U , are calculated on the assumption that the crust is a closed system and incorporated U free of Th and Pa at the time of formation. To lower the contamination by Th from hydrogenous or sedimentary components, we have selected samples with compact manganese phases represented by very low ^{232}Th contents. This sedimentary component, always important in the outermost layer, thus precludes the dating of the assumed youngest layer. Any contamination would lead to older ages and therefore the ages given are probably overestimated.

Sample DR11/B, identified as a hydrogenous incrustation from its chemical and mineralogical composition, has a high ^{232}Th content and the large excess of ^{230}Th and ^{231}Pa normally encountered in such deposits⁶. The $^{234}\text{U}/^{238}\text{U}$ ratio does not differ significantly from the seawater value of 1.14 ± 0.03 ¹³. With a ^{210}Pb content of 770 ± 70 d.p.m. g⁻¹ and a ^{226}Ra content of 41 ± 4 d.p.m. g⁻¹, we calculated a $^{210}\text{Pb}/^{226}\text{Ra}$ ratio of 19 and a $^{226}\text{Ra}/^{230}\text{Th}$ ratio of 0.15. This excess of ^{210}Pb and the deficiency

of ^{226}Ra confirm that the crust was exposed to seawater at the time of sampling¹⁴. The nontronite sample has very low U and Th activities. This may result from contamination either by detrital sediment or by Mn as suggested by its Al and Mn content. Sampling of pure nontronite is impossible as it occurs only as thin lenses in the crusts, but our measurements confirm the strong fractionation of the elements during the hydrothermal process, with U being associated with the Mn oxides rather than with the Fe–Si precipitate.

The most striking feature in Table 2 is the high $^{234}\text{U}/^{238}\text{U}$ ratio measured in all the Mn rich crusts. The ratio of 2.25 ± 0.03 , about twice the seawater ratio, is the highest so far measured in an oceanic material. There are two indications that no post-depositional migration, as described in other deep sea deposits,^{4,15} occurred in our samples: (1) in the crust DR4/B the U content of the deep layer (45–50 mm) is twice that of the shallower layer (11–23 mm) but the $^{234}\text{U}/^{238}\text{U}$ is identical, (2) the ages obtained from the ingrowth of ^{230}Th and ^{231}Pa to their parents ^{234}U and ^{235}U are in good agreement. Extensive mobility of the U isotopes within the crust that would change the ratio by a factor of two would lead to discordant ages by both methods.

The U isotopes found in the Mn layers have four possible sources: seawater, pore water, hydrothermal solutions and detrital sediments included in the oxides. As the U content of the crust is higher than that of the sediment, this last source can be safely rejected. There is no known mechanism by which the U isotopes of seawater could be fractionated during deposition; indeed the hydrogenous deposits display a typical seawater $^{234}\text{U}/^{238}\text{U}$ ratio, so that this source can also be discounted. All published radiochemical data on diagenetic Mn crusts^{12,16–18} as well as direct measurements in pore waters¹⁹ are close to the seawater $^{234}\text{U}/^{238}\text{U}$ ratio. This leaves a hydrothermal solution as the only possible U source with very high isotopic ratio.

In nature, U isotopes are fractionated by the effects of alpha recoil. Differences in bonding and oxidation state render the ^{234}U daughter isotope more mobile than its ^{238}U parent⁸. The anomalously high $^{234}\text{U}/^{238}\text{U}$ reported in hydrothermal Fe–Mn crusts^{1–5} and in some late-stage volcanic products²⁰ have been attributed to such 'hot atom' effects. Krishnaswami and Turekian²¹ and Michard *et al.*²² measured U in hydrothermal vent solutions. They found no U in high-temperature fluids and deduced²¹ that this element was deposited at temperatures $\geq 30^\circ\text{C}$. Following Rydell and Bonatti², they infer that seawater entering the system should be stripped of its U and deposited in hydrothermal conduits. Leaching of U from the rocks during shallow stages of sub-bottom circulation at lower temperatures may produce a solution relatively enriched in ^{234}U . The mixing of the hydrothermal solution with oxidizing seawater would result in variable $^{234}\text{U}/^{238}\text{U}$ ratios, ranging from the seawater ratio to higher values. In our samples, however, the ratios do not seem to be related to the ^{238}U content. Thus, a simple mixing process is not convincing. If the seawater U characteristics are known to be constant, the ^{234}U enrichment of the hydrothermal

solution should depend on the residence time of the fluid in the rocks.

Kazachevskii *et al.*²³ measured a $^{234}\text{U}/^{238}\text{U}$ ratio of 1.14 ± 0.02 in Fe-Mn suspensions sampled directly on the flank of the volcano Banu Vukhu in the Sanghihe archipelago, the same arc system as our sampling area, where the emanation interacts with seawater. As the precipitation was observed to take place in seawater, it is likely that the hydrothermal U contribution in this sample was masked by the seawater imprint, as observed in other hydrothermal oxide deposits. On board, 3.5-kHz echo sounder records suggest that the crusts examined lay on a thin sedimentary cover. We hypothesize that the precipitation of the hydrothermal solution does take place in/at the top of this sediment before it is mixed with seawater. The origin of the U measured in the sediments from the Galapagos Mounds by Lalou and Brichet²⁴ may be due to a similar process as described by Moorbey²⁵. Several authors^{26,27} have found evidence of a U source with $^{234}\text{U}/^{238}\text{U} > 1.14$, the extension of the present data to other low-temperature hydrothermal systems might clarify the hydrothermal solutions as the source involved. With the lack of direct measurements of the hydrothermal solution itself, we suggest that the anomalous $^{234}\text{U}/^{238}\text{U}$ ratio in seafloor deposits provide a probe for the low temperature hydrothermal reactions in the ocean crust and thereby could be used, together with ancillary chemical information, to study chemical processes occurring in such systems. Work in progress in our laboratory is designed to contribute to this aim.

We thank Dr L. Labeyrie, chief scientist of the cruise ESTASE and the Captain and crew of RV *Charcot*, C. Lalou, S. Calvert, T. Clube and J. C. Duplessy for helpful comments, and J. Devineau for technical assistance. The manuscript was improved

by critical reviews from A. Aplin and an anonymous reviewer. Laboratory studies were funded by the C.E.A. and ATP-CNRS Océanographie Chimique (CFR contribution no. 800).

Received 5 August; accepted 3 December 1986.

1. Veeh, H. H. & Boström, K. *Earth planet. Sci. Lett.* **10**, 372-374 (1971).
2. Rydell, H. & Bonatti, E. *Geochim. cosmochim. Acta* **37**, 2557-2565 (1973).
3. Piper, D. Z., Veeh, H. H., Bertrand, W. G. & Chase, R. L. *Earth planet. Sci. Lett.* **26**, 114-120 (1975).
4. Moore, W. S. & Vogt, P. R. *Earth planet. Sci. Lett.* **29**, 349-356 (1976).
5. Lalou, C., Brichet, E., Jehanno, C. & Leclaire, H. *Earth planet. Sci. Lett.* **63**, 63-75 (1983).
6. Ku, T. L. & Broecker, W. S. *Deep Sea Res.* **16**, 625-637 (1969).
7. Cronan, D., Glasby, G. P., Moorbey, S. A., Thompson, J., Knedler, K. E. & McDougall, J. C. *Nature* **298**, 456-458 (1982).
8. Ivanovich, M. & Harmon, R. S. *Uranium Series Disequilibrium: Applications to Environmental Problems* (Clarendon, Oxford, 1982).
9. Moorbey, S. A., Cronan, D. S. & Glasby, G. P. *Geochim. cosmochim. Acta* **48**, 433-441 (1984).
10. Corliss, J. B., Lyle, M., Dymond, J. & Crane, K. *Earth planet. Sci. Lett.* **40**, 12-24 (1978).
11. Dymond, J., Lyle, M., Finney, B., Piper, D. Z., Murphy, K., Conrad, R. & Pisias, N. *Geochim. cosmochim. Acta* **48**, 931-949 (1984).
12. Reyss, J. L., Marchig, V. & Ku, T. L. *Nature* **295**, 401-403 (1982).
13. Ku, T. L., Knauss, K. G. & Mathieu, G. G. *Deep Sea Res.* **24**, 1005-1007 (1977).
14. Krishnaswami, S. & Cochran, J. K. *Earth planet. Sci. Lett.* **40**, 45-62 (1978).
15. Kolodny, Y. & Kaplan, I. R. *Geochim. cosmochim. Acta* **34**, 3-24 (1970).
16. Ku, T. L. & Glasby, G. P. *Geochim. cosmochim. Acta* **36**, 699-703 (1972).
17. Huh, C. A. & Ku, T. L. *Geochim. cosmochim. Acta* **48**, 951-963 (1984).
18. Reyss, J. L., Lemaire, N., Ku, T. L., Marchig, V., Southon, J. R., Nelson, J. R. & Vogel, J. S. *Geochim. cosmochim. Acta* **49**, 2401-2408 (1985).
19. Cochran, J. K. & Krishnaswami, S. *Am. J. Sci.* **280**, 849-899 (1980).
20. Cherdynstev, V. V., Kupstiv, V. M., Kuz'mina, E. A. & Zverev, V. L. *Geokhimiya* **1**, 77-85 (1968).
21. Krishnaswami, S. & Turekian, K. K. *Geophys. Res. Lett.* **9**, 827-830 (1982).
22. Michard, A., Albareda, F., Michard, G., Minster, J. F. & Charlou, J. L. *Nature* **303**, 795-797 (1983).
23. Kazachevskii, I. V., Cherdynstev, V. V., Kuz'mina, E. A., Sulerzhitskii, L. D., Mochalova, V. F. & Kyuregyan, T. N. *Geokhimiya* **11**, 1116-1121 (1964).
24. Lalou, C. & Brichet, E. *Nature* **284**, 251-253 (1980).
25. Moorbey, S. A. *Earth planet. Sci. Lett.* **62**, 367-376 (1983).
26. MacDougall, J. C., Finkel, R. C., Carlson, J. & Krishnaswami, S. *Earth planet. Sci. Lett.* **42**, 27-34 (1979).
27. Borole, D. V., Krishnaswami, S. & Somayajulu, B. L. K. *Geochim. cosmochim. Acta* **46**, 125-137 (1982).

Long-range transport of terrestrially derived lipids in aerosols from the south Pacific

Robert B. Gagosian*, Edward T. Peltzer* & John T. Merrill†

* Department of Chemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

† Center for Atmospheric Chemistry Studies, Graduate School of Oceanography, University of Rhode Island, Kingston, Rhode Island 02881, USA

Terrestrially produced particulate organic matter can be transported relatively fast in the atmosphere, long distances over the ocean. It has been hypothesized that this atmospheric transport may be an important way of quickly introducing continentally-derived organic material to the surface waters of the open ocean¹⁻¹¹. After rapid transport through the water column to the sediment surface¹², these terrestrial organic substances could account for an important fraction of the organic carbon found in the sedimentary record¹³. The atmospheric fluxes of these organic substances are large enough to have a major potential impact on the inventory of terrestrially derived lipid material, originating from vascular plant waxes, found in deep-sea sediments¹³. We present here the first use of organic compound biological source marker information in conjunction with long-range meteorological trajectory analysis to elucidate specific terrestrial source regions and to determine the transport pathways of organic material through the atmosphere to remote marine locations.

The samples were collected as part of the south Pacific portion of the Sea/Air Exchange Program (SEAREX) in May-September 1983, from a 20-m tower erected on Ninety Mile Beach on the northwestern shore of the North Island of New Zealand. The site, 34.6° S, 172.8° W, is located in the Southern

Hemisphere westerlies wind regime, approximately 1,500 km east of Australia. The sampling-analysis scheme for lipids in aerosols is summarized by Gagosian *et al.*⁸ and the detailed analytical methods are presented in Peltzer *et al.*¹⁴. This sampling-analysis process requires stringent precautions against contamination during sampling, storage, transport, and analysis due to the extremely low concentrations of organic compounds found in remote marine air. The analysis involved extraction of glass fibre filter samples with methylene chloride, separation into lipid compound classes by silica-gel column chromatography and quantification by glass-capillary gas chromatography. Compound identifications were verified by gas-chromatography mass spectrometry and comparison of these spectra with authentic standards. Samples were accompanied by appropriate tower-to-analysis blanks. Recovery factors were determined by spiking the samples with appropriate internal standards¹⁴.

To identify the sources of organic substances in aerosols, we have used the biomarker approach which takes advantage of the unique molecular signatures or source 'fingerprints' for various lipid compound classes of marine and terrestrial plants and animals and of anthropogenic activities¹¹. Although these compounds make up approximately 10-50% of the solvent extractable organic matter in aerosols⁸, they represent only a small portion of the total organic material. This 'tracer' approach for lipid class compounds is analogous to the use of natural and bomb produced radionuclides in atmospheric and oceanic source and transport studies.

Several conclusions concerning the sources of the organic material associated with aerosols can be made by examining the individual compound distributions within specific lipid compound classes. For all the samples analysed, the major source signature for the hydrocarbons was from terrestrial plant waxes^{1,15,16} (Table 1, Fig. 1). For the *n*-alkanes, a strong odd-even carbon preference was observed, the major homologues being C₂₉ and C₃₁. However, for the first time we also observed the lighter *n*-alkanes in the range C₁₇-C₂₀. This is most likely due to the lower ambient temperatures in New Zealand com-

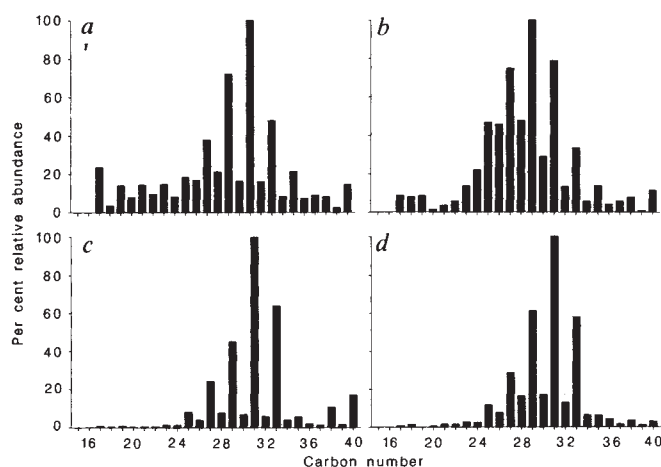
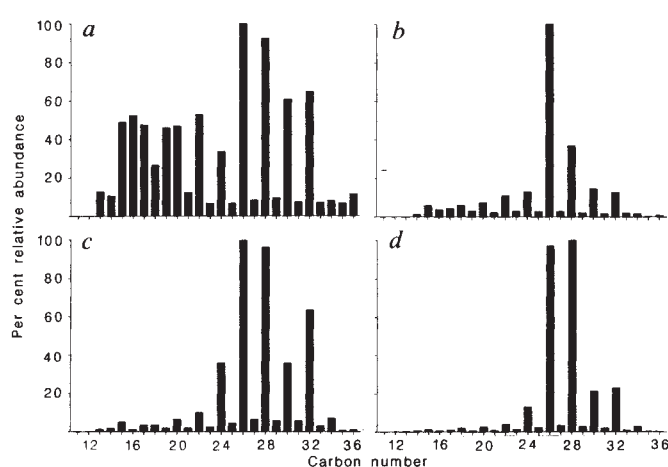
Table 1 Atmospheric aerosol concentrations of aliphatic hydrocarbon, fatty alcohols, and fatty acid salts from samples collected from Ninety Mile Beach, New Zealand

Sample no. (mid-date/volume) (m ²)	NZAS-5 (19 June/4,000 m ³)	NZAS-16 (18 July/3,700 m ³)	NZAS-20 (12 Aug/4,400 m ³)	NZAS-22 (21 Aug/3,000 m ³)
Aliphatic hydrocarbons				
C ₁₇ -C ₄₀ *	22	61	186	270
CPI† (C ₁₇ -C ₄₀)	3.5	2.4	8.5	4.3
Major components‡	1, 29, 33	29, 31, 27	31, 33, 29	31, 29, 33
Fatty alcohols				
C ₁₃ -C ₂₀ *	16	10	16	14
Major components‡	16, 15, 17	20, 18, 15	20, 15, 18	20, 19, 18
C ₂₁ -C ₃₆ *	32	76	291	474
CPI† (C ₂₁ -C ₃₆)	9.2	14.4	13.3	22.1
Major components‡	26, 28, 32	26, 28, 30	26, 28, 32	28, 26, 32
Fatty acid salts				
C ₁₃ -C ₁₈ *	565	636	623	333
Major components‡	16, 14, 18	16, 14, 18	16, 14, 18	16, 18, 14
C ₁₉ -C ₃₆ *	65	86	81	98
CPI† (C ₁₉ -C ₃₆)	2.5	2.4	2.2	2.2
Major components‡	24, 20, 22	22, 24, 20	20, 24, 22	20, 22, 24

* Concentrations are in pg m⁻³ air.

† Carbon number in order of decreasing abundance.

‡ CPI, carbon preference index. Odd-even for aliphatic hydrocarbons, even-odd for fatty alcohols and acids.

**Fig. 1** Aliphatic hydrocarbon distribution patterns for aerosols collected at Ninety Mile Beach, New Zealand, June–August 1983. *a*, NZAS-5; *b*, NZAS-16; *c*, NZAS-20; *d*, NZAS-22. Total hydrocarbon concentrations are 22, 61, 186 and 270 pg m⁻³ respectively.**Fig. 2** Fatty alcohol distribution patterns for aerosols collected at Ninety Mile Beach, New Zealand, June–August 1983. *a*, NZAS-5; *b*, NZAS-16; *c*, NZAS-20; *d*, NZAS-22. Total fatty alcohol concentrations are 48, 86, 307 and 488 pg m⁻³ respectively.

pared to the higher tropical temperatures found at earlier sampling sites (American Samoa in the south Pacific¹⁷ and Enewetak Atoll in the north Pacific⁸). For in-sector (ISS) westerly wind samples (NZAS-5, 20 and 22), which all had an Australian continental origin, C₃₁ was the major homologue, where as C₂₉ was the major component of the *n*-alkanes in aerosols coming from across New Zealand (NZAS-16). For many of the samples that we have analysed from American Samoa, C₃₁ was the major *n*-alkane. This suggests that tropical *n*-alkane aerosol sources have higher carbon number homologues than their temperate counterparts¹⁷. The lower molecular weight source signature for NZAS-16 is consistent with its more temperate source.

The fatty alcohol distributions add further evidence to these source region assignments (Table 1, Fig. 2). In all the samples, a strong even-odd carbon preference for the C₂₁-C₃₆ range was observed which is an indicator of terrestrial plant wax input^{15,16}. Concentrations of the low molecular weight fatty alcohols (C₁₄-C₂₀) were low (10–20 pg m⁻³) and fairly constant with time as expected for a marine source (Table 1). On the other hand, the terrestrially derived fatty alcohols (C₂₁-C₃₆) vary in concentra-

tion by over an order of magnitude. For the New Zealand ISS samples (NZAS-5, 20 and 22), the fatty alcohol mean carbon number was 27.5–28. At Enewetak the modal carbon number for the fatty alcohols showed a steady temporal progression from 27.1 to 29.3¹³. Based on long-range meteorological transport data¹⁹, it was concluded that this progression was indicative of a shift from a temperate source in central Asia to a more tropical Central American source later in the year. In contrast to the three New Zealand ISS samples, the mean carbon number of the NZAS-16 fatty alcohols was 26.8. In addition, unlike most other samples where C₂₆ and C₂₈ are nearly equal in abundance, in NZAS-16 the C₂₆ fatty alcohol was 42% of the total fatty alcohols and was three times more abundant than C₂₈. This indicates that NZAS-16 has a substantially different and more temperate source than the three ISS samples. Sample NZAS-15 which was taken one week before NZAS-16, has the same long range trajectory pattern and exhibits a similar fatty alcohol distribution, C₂₆ being 58.5% of the total fatty alcohol concentration.

The low molecular weight (C₁₃-C₁₈) fatty acid salts appear

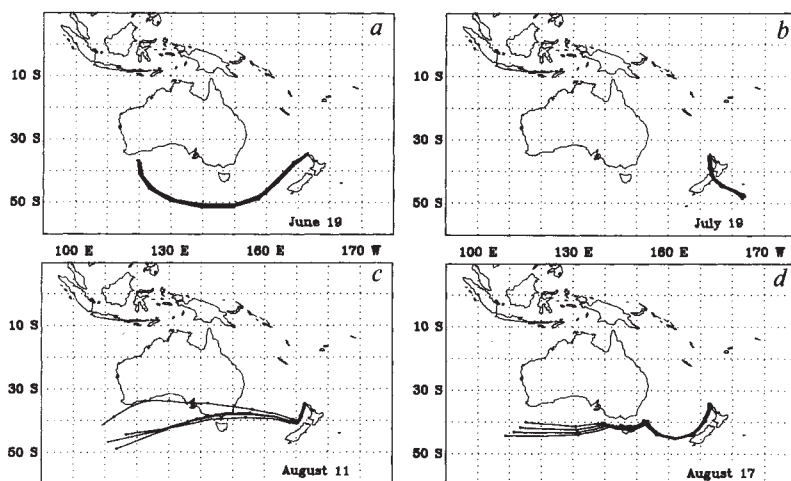


Fig. 3 Isentropic trajectories backward in time from the experiment site. Examples shown are at 290 K and are in the boundary layer at the site. They are representative of ensembles of trajectories calculated at surrounding points¹⁹. The dots on the trajectories are at one day intervals.

to have a marine origin¹⁵. In all of the New Zealand samples, C_{16} was the most abundant homologue followed by C_{14} and C_{18} (Table 1). This is comparable to the abundance patterns seen in the Enewetak⁸ and Samoa¹⁷ samples. The higher molecular weight (C_{19} – C_{32}) fatty acid salts in contrast show a pronounced land plant source as evidenced by the strong even-odd carbon predominance pattern and the lack of any real major homologue in the range C_{21} – C_{32} ¹⁵. The mean carbon number of 24 ± 0.5 is typical of this continental source as well.

Further information can be obtained concerning the source of the atmospheric aerosols by examining the New Zealand aerosol data in conjunction with the long-range transport and air parcel trajectory meteorological data. Figure 1 shows the aliphatic hydrocarbon concentrations for four New Zealand air samples. NZAS-5 had the lowest concentrations of all the New Zealand samples, 22 pg m^{-3} (Fig. 1a). The *n*-alkane composition for this sample had a lower mean carbon number than either NZAS-20 or NZAS-22 suggesting a more temperate source. The fatty alcohol composition of NZAS-5 (Fig. 2a) suggested a mixture of several sources including tropical, temperate and marine. Thus, based on chemical evidence we conclude that this sample represents southern ocean air with some minor continental influence.

NZAS-16 (Fig. 1b) had moderate levels of all three compound classes suggesting a stronger continental source than NZAS-5 but not nearly as strong as NZAS-20 or NZAS-22. The presence of anthropogenic *n*-alkanes in the hydrocarbon fraction (low CPI (carbon preference index), Table 1) is indicative of an urban source while the very low mean carbon number for the high molecular weight fatty alcohols (Fig. 2b) suggests that this sample had the most temperate source of all the samples analysed. Thus, we would predict that this sample represents an urban source in New Zealand. However, the *n*-alkane and terrestrial fatty alcohol concentrations were quite low (61 and 76 pg m^{-3} respectively) suggesting either that the transport was from high level winds over the source or that the source was quite distant.

Samples NZAS-20 and NZAS-22 had very high and similar concentrations of *n*-alkanes (186 and 270 pg m^{-3}) and terrestrial fatty alcohols (291 and 474 pg m^{-3}). The *n*-alkanes, the high molecular weight fatty alcohols and the high molecular weight fatty acid salts all indicate a major terrestrial source. Both samples exhibited a more tropical character than NZAS-16, NZAS-20 being the more tropical as exemplified by the high concentration of the C_{32} fatty alcohol. The levels of *n*-alkanes and fatty alcohols in NZAS-20 and -22 were greater than in any of our previous samples^{8,11,17} suggesting a close continental source, most probably Australia.

In Fig. 3, we present air mass trajectories which are typical of the trajectories followed by the air samples in New Zealand for each of the aerosol samples. The trajectories were calculated using wind fields on isentropic surfaces found by the process described by Merrill *et al.*^{18,19}. In each case, these trajectories are consistent with the predicted source of the aerosols based on the trace organic compound analyses. Figure 3a clearly shows a southern ocean origin for NZAS-5 which is essentially free of all continental influence for many thousands of miles. Some trajectories, during the period of sampling, approach the far southwestern part of the South Island of New Zealand. Most, however, are similar to the case shown, where the 'origin' (the position 10 days before arrival at the sampling site), is in the open ocean or near southwestern Australia. Figure 3b shows that NZAS-16 contains aerosols from New Zealand and that there was urban and terrestrial influence from the south of the North Island and Christchurch. The relatively low input of anthropogenic material found on the aerosol sample is supported by the trajectory analyses which show that the winds passed over New Zealand at heights above the boundary layer (800 – 900 mb). The trajectories for aerosol samples NZAS-20 (Fig. 3c) and NZAS-22 (Fig. 3d) originated in southern Australia and Tasmania and are consistent with the strong terrestrial biomarker signal. These examples of correlations of trace organic compound and air parcel trajectory analyses clearly show the benefits of this approach in defining aerosol sources and transport pathways. Future studies will help us further quantify source regions and place boundary conditions on the meteorological models.

We thank the Oceanography and Atmospheric Sciences Sections of the NSF for support as part of the SEAREX programme. The meteorological calculations were performed with computer facilities at the National Center for Atmospheric Research, also supported by the NSF. We thank J. B. Alford and N. M. Frew for assistance in the analytical work and The New Zealand Meteorological Service and the Te Pahi Farm Park Ranger for their assistance in the field.

Received 16 September; accepted 11 December 1986.

1. Simoneit, B. R. T. *Mar. Chem.* **5**, 443–464 (1977).
2. Simoneit, B. R. T. in *Proceedings Carbonaceous Particles in the Atmosphere* 233–244 (ed. Novokov, T.) NSF-LBL (1979).
3. Eichmann, R., Neuling, P., Ketscherdis, G., Hahn, J., Jaenicke, R. & Jung, C. *Atmos. Envir.* **13**, 587–599 (1979).
4. Eichmann, R. *et al. Atmos. Envir.* **14**, 695–703 (1980).
5. Gagosian, G. B., Peltzer, E. T. & Zafiriou, O. C. *Nature* **291**, 312–314 (1981).
6. Chesselet, R., Fontugne, M., Buat-Menard, P., Ezat, U. & Lambert, C. E. *Geophys. Res. Lett.* **8**, 345–348 (1981).

7. Marty, J. C. & Sillot, A. *Nature* **298**, 144–147 (1982).
8. Gagosian, R. B., Zafriou, O. C., Peltzer, E. T. & Alford, J. B. *J. geophys. Res.* **87**(C), 11133–11144 (1982).
9. Ohta, K. & Handa, N. *J. Ocean. Soc. Jpn.* **41**, 25–32 (1985).
10. Zafriou, O. C., Gagosian, R. B., Peltzer, E. T. & Alford, J. B. *J. geophys. Res.* **90**, 2409–2423 (1985).
11. Gagosian, R. B. in *NATO Advanced Study Institute on the Role of Air/Sea Exchange in Geochemical Cycling* (ed Buat-Ménard, P.) 409–442 (Reidel, Dordrecht, 1986).
12. Wakeham, S. G., Lee, C., Farrington, J. W. & Gagosian, R. B. *Deep-Sea Res.* **31**, 509–528 (1984).
13. Gagosian, R. B. & Peltzer, E. T. *Org. Geochem.* (in the press).
14. Peltzer, E. T., Alford, J. B. & Gagosian, R. B. *Tech. Rep.* 84–89 (Woods Hole Oceanographic Institution, Woods Hole, 1984).
15. Kolattukudy, P. E. *Chemistry and Biochemistry of Natural Waxes* (Elsevier, New York, 1976).
16. Eglinton, G., Gozalez, A. G., Hamilton, R. J. & Raphael, R. A. *Photochemistry* **1**, 89–102 (1962).
17. Peltzer, E. T. & Gagosian, R. B. *SEAREX News.* **7**(1), 22–26 (1984).
18. Merrill, J. T., Bleck, R. & Avila, L. *J. geophys. Res.* **90**, 12927–12936 (1985).
19. Merrill, J. T., Bleck, R. & Boudra, D. *Mon. Weath. Rev.* **114**, 571–581 (1986).

Primary production, new production and vertical flux in the eastern Pacific Ocean

Michael L. Pace*, George A. Knauer†, David M. Karl‡ & John H. Martin§

* Institute of Ecosystem Studies, The New York Botanical Garden, Mary Flagler Cary Arboretum, Box AB, Millbrook, New York 12545 USA

† Center for Marine Science, University of Southern Mississippi, National Space Technology Laboratories, NSTL, Mississippi 39529, USA

‡ Department of Oceanography and Hawaii Institute of Geophysics, University of Hawaii, 1000 Pope Road, Honolulu, Hawaii 96822, USA

§ Moss Landing Marine Laboratories, Moss Landing, California 95039, USA

The sinking of particulate organic matter in the ocean links food webs beneath the euphotic zone to surface primary production and is an important pathway for the downward transport of many elements^{1–3}. The flux of particulate organic carbon (POC) is also an important parameter in the global carbon cycle and may be related to long-term changes in atmospheric CO₂^{4,5}. In 1980, Suess⁶ synthesized existing measurements from sediment trap studies into a model to predict the vertical flux of POC from depth (*z*) and primary production (PP)⁶. The Suess model has become the standard for evaluating vertical flux data⁷, for estimating the annual flux of POC in the ocean⁸ and for parameterizing ocean carbon cycle models^{4,5}. We present here a new model of the vertical flux of POC and particulate organic nitrogen (PON) from a set of contemporaneous measurements of PP and fluxes made during the VERTEX (Vertical Transport and Exchange) programme in the north-east Pacific. The VERTEX model indicates that PP and vertical fluxes of POC and PON, in the oligotrophic ocean are greater than previously suggested. In addition, the vertical flux of PON from the photic zone represents a measure of the PP that is supported by new nitrogen (new production)^{9,10}. In the north-east Pacific, new production ranged from 13 to 25% of primary production and was positively related to total PP.

The vertical flux of POC and PON in the upper 2,000 m of the water column was measured using free floating particle

traps¹¹ at nine sites in the Pacific Ocean^{12,13}. While traps were deployed, PP was measured at each site, using the heavy-metal-free ¹⁴C method (ref. 14). Trapped materials were preserved, in a formalin solution¹⁵, as collected to avoid losses due to decomposition. Some organics undoubtedly partition from the particulate to dissolved fraction during preservation so the particulate fluxes measured in sediment traps represent conservative estimates^{15,16}. All collections were also examined microscopically and zooplankton which had entered the traps alive during the deployment were removed before measurement of POC and PON¹⁵.

The VERTEX data were log₁₀-transformed and relationships for POC and PON were derived using multiple regression with *z* and PP as independent variables. We chose to express the data in terms of mg m⁻² d⁻¹ instead of the units of g m⁻² yr⁻¹ used in previous models to reflect more accurately the timescale of the measurements. The following equations for POC and PON were obtained:

$$\text{POC} = 3.523 z^{-0.734} \text{PP}^{1.000} \quad (1)$$

$$\text{PON} = 0.432 z^{-0.843} \text{PP}^{1.123} \quad (2)$$

where *z* is restricted to depths between the bottom of the euphotic zone and 2,000 m.

For comparison, we recalculated in terms of daily flux the original data from the model of Suess⁶ and a reformulation of a limited portion of the same data by Betzer *et al.*⁷ (Table 1). All models are highly significant; however, the amount of variance explained (*r*²) in the VERTEX models is lower than previous models (Table 1). We attribute this partly to the higher variability that might be expected in contemporaneous measures of PP and flux relative to the averages of historical data used in previous models. In addition, increased flux rates with depth were occasionally observed, implying that horizontal advection of particles and *in situ* particle production may occur at specific depths^{12,17,18}.

The partial regression coefficients indicate the form of the relationship between vertical flux and the independent variables. For all three POC regressions, the partial regression coefficients for *z* are significantly < -1 (*t*-test, all *P* < 0.001). The nonlinear change in the relationship between flux and depth indicates, not unexpectedly¹⁹, that most particulates are transformed to biomass, dissolved organics, and inorganic end products in the

Table 1 Summary of multiple regression models to predict the vertical flux of particulate organic carbon (POC) and nitrogen (PON) in mg C or N m⁻² d⁻¹ from depth (*z*) in metres and primary production (PP) in mg C m⁻² d⁻¹

Source	<i>n</i>	<i>B</i> ₀	<i>B</i> ₁	±95% CI	Range	$\bar{Z}$	<i>B</i> ₂	±95% CI	Range	$\bar{PP}$	<i>r</i> ²	F	CF
Suess													
POC	33	-0.503	-0.567	0.104	50–5,582	780	1.302	0.164	110–1,644	345	0.97	507	1.066
Betzer <i>et al.</i>													
POC	27	-0.574	-0.627	0.121	100–5,582	1,392	1.415	0.151	110–1,644	272	0.97	433	1.043
VERTEX													
POC	74	0.547	-0.734	0.143	50–2,000	497	1.000	0.357	245–1,140	517	0.69	79.3	1.221
VERTEX													
PON	74	-0.364	-0.843	0.139	50–2,000	497	1.123	0.345	245–1,140	517	0.76	114	1.205

Data were log₁₀ transformed and fitted to the equation log₁₀ Flux = *B*₀ + *B*₁ log₁₀ *z* + *B*₂ log₁₀ PP. Additional statistics are the means and ranges of the data used in each model as well as the coefficient of determination (*r*²) and F-value from the multiple regression analysis of variance (for all 4 models *P* < 0.0001). Predictions in arithmetic units from these logarithmic regressions must be corrected for bias introduced in back transformation using the calculated correction factor (CF)²⁷.

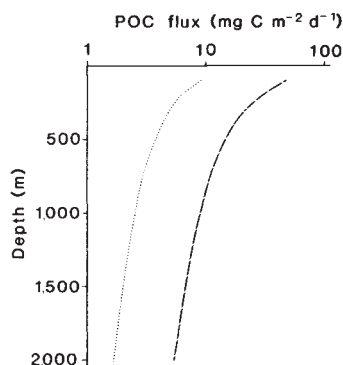


Fig. 1 Comparison of the Suess (dotted line) and VERTEX (dashed line) models at primary productivities representative of oligotrophic oceanic conditions. For the Suess model a historical average PP of $100 \text{ mg C m}^{-2} \text{ d}^{-1}$ was used. For the VERTEX model the median measured oceanic PP of $400 \text{ mg C m}^{-2} \text{ d}^{-1}$ was used.

upper water column. The partial regression coefficients for PP are typically >1 , but not significantly so for the VERTEX models. The hypothesis that the ratio of flux to production increases with PP ($B_2 > 1$) (see ref. 7) is not supported by the VERTEX data.

The PP data used in the Suess model reflect historical measurements and are lower than measurements made in the VERTEX programme. For example, daily productivities of the order of $100\text{--}150 \text{ mg C m}^{-2} \text{ d}^{-1}$ were considered representative of oligotrophic oceanic conditions⁶. Measurements of PP made at 5 oligotrophic sites during VERTEX ranged from $245\text{--}760 \text{ mg C m}^{-2} \text{ d}^{-1}$ with a median value of $400 \text{ mg C m}^{-2} \text{ d}^{-1}$. Figure 1 compares predicted POC fluxes for the Suess and VERTEX models at PP of 100 and $400 \text{ mg C m}^{-2} \text{ d}^{-1}$ respectively. The VERTEX model POC fluxes are 3–5 times greater (depending on depth) under these conditions. The actual rates of PP in the open ocean have recently been intensely debated^{20–25}. The VERTEX data, although limited, support the notion that oceanic PP is higher than historical estimates and that the vertical flux of sinking particles is also greater.

The flux of PON from the euphotic zone represents a measure of new production, since this flux must ultimately be balanced by inputs of nitrogen to the euphotic zone^{9,10}. For the nine study sites, the relationship between PP and the flux of PON is significant (Fig. 2). Using the linear regression model of Fig. 2 and assuming a C/N ratio of 6.6, new production as a percentage of the total increases from 13 to 25% over the PP range of $245\text{--}1,140 \text{ mg C m}^{-2} \text{ d}^{-1}$. Confidence, however, in any prediction of new production from measures of PP and our model is poor (Fig. 2). This might be interpreted as support for the notion of high variability in new production²⁴, but for most oligo-

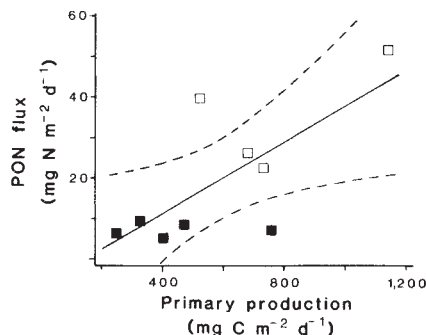


Fig. 2 Relationship between primary production and vertical flux of particulate organic nitrogen (PON) measured in traps located at the base of the euphotic zone for nine VERTEX stations. Regression model: $\text{PON} = 0.0439 \text{ PP} - 6.17$, $r^2 = 0.52$, $F = 7.57$, $P < 0.01$. Solid squares represent stations with deep euphotic zones ($\geq 100 \text{ m}$) and open squares represent stations with euphotic zones of 50 m . Broken lines are the 95% confidence intervals for the regression.

trophic sites studied with euphotic zone depths $\geq 100 \text{ m}$ ($n = 5$, solid squares, Fig. 2), new production as measured by PON flux was relatively constant: $\bar{x} \pm 1 \text{ s.d.} = 7.23 \pm 1.59 \text{ mg N m}^{-2} \text{ d}^{-1}$ (Fig. 2). If we convert the daily mean new production value for oligotrophic conditions to an annual rate by multiplying by 365 and assuming a Redfield C/N ratio of 6.6, the estimate of $17.4 \pm 3.83 \text{ g C m}^{-2} \text{ yr}^{-1}$ is reasonably close to the value of $25.2 \pm 6.96 \text{ g C m}^{-2} \text{ yr}^{-1}$ recently estimated for the Sargasso Sea²³. Our value, however, is significantly less than the $50 \text{ g C m}^{-2} \text{ yr}^{-1}$ estimated by Jenkins and Goldman²³. The data are clearly too limited to provide an accurate estimate of new production, given the potential spatial and temporal variability. A VERTEX seasonal study to measure the variability in PP and PON flux over an annual cycle at an oceanic site is now underway.

The empirical models of vertical flux allow prediction of the rate of loss of organic matter from the photic zone, the transformation of this material in the mesopelagic zone, and input to the deep sea. We conclude that the magnitude of vertical flux in the ocean may be greater than previous models based on PP suggest, because oceanic PP may be higher than historical averages. The coupling of models of vertical flux as developed in the VERTEX study with improved models of circulation and PP based on remote sensing will substantially enhance our knowledge of the flux and cycling of materials on seasonal, regional and basin-wide scales in the ocean²⁶.

We thank C. Hunter and M. Tuel for help in the field and laboratory and N. Caraco and J. Cole for comments on the manuscript. This research was supported by grants from the NSF Marine Chemistry and Biological Oceanography programmes.

Received 14 July; accepted 10 December 1986.

1. Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Columbia University, Palisades, New York, 1982).
2. Angel, M. V. in *Flows of Materials and Energy in Marine Ecosystems* (ed. Fasham, M. J. R.) 475–516 (Plenum, New York, 1984).
3. Fowler, S. W. & Knauer, G. A. *Prog. Oceanogr.* **16**, 147–194 (1986).
4. Sarmiento, J. L. & Toggweiler, J. R. *Nature* **308**, 621–624 (1984).
5. Dymond, J. & Lyle, M. *Limnol. Oceanogr.* **30**, 699–712 (1985).
6. Suess, E. *Nature* **288**, 260–263 (1980).
7. Betzer, P. R. *et al. Deep Sea Res.* **31**, 1–11 (1984).
8. Hargrave, B. T. *Ecol. Modell.* **30**, 229–246 (1985).
9. Dugdale, R. C. & Goering, J. J. *Limnol. Oceanogr.* **12**, 196–206 (1967).
10. Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677–680 (1979).
11. Knauer, G. A., Martin, J. H. & Bruland, K. W. *Deep Sea Res.* **26**, 97–108 (1979).
12. Knauer, G. A., Martin, J. H. & Karl, D. M. in *Global Ocean Flux Study Proc. Workshop* 136–150 (National Academy Press, Washington DC, 1984).
13. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. *Deep Sea Res.* (in the press).

14. Fitzwater, S. E., Knauer, G. A. & Martin, J. H. *Limnol. Oceanogr.* **27**, 544–551 (1982).
15. Knauer, G. A., Karl, D. M., Martin, J. H. & Hunter, C. N. *J. mar. Res.* **42**, 445–462 (1984).
16. Knauer, G. A. & Martin, J. H. in *Second US-USSR Symposium on Biological Effects of Pollutants on Marine Organisms* 145–165 (EPA-600/3-82-034, Washington DC, 1982).
17. Karl, D. M., Knauer, G. A., Martin, J. H. & Ward, B. B. *Nature* **309**, 54–56 (1984).
18. Urrère, M. A. & Knauer, G. A. *J. plank. Res.* **3**, 369–387 (1981).
19. Menzel, D. W. & Ryther, J. D. in *Organic Matter in Natural Waters* (ed. Hood, D. W.) 31–54 (University of Alaska, 1970).
20. Shulenberger, E. & Reid, J. L. *Deep Sea Res.* **28**, 901–919 (1981).
21. Jenkins, W. J. *Nature* **300**, 246–248 (1982).
22. Platt, T. *Deep Sea Res.* **31**, 1311–1319 (1984).
23. Jenkins, W. J. & Goldman, J. C. *J. mar. Res.* **43**, 465–491 (1985).
24. Platt, T. & Harrison, W. G. *Nature* **318**, 55–58 (1986).
25. Kerr, R. A. *Science* **232**, 1345 (1986).
26. National Academy of Sciences, *Global Ocean Flux Study Proc. Workshop* (National Academy Press, Washington DC, 1984).
27. Sprugel, D. G. *Ecology* **64**, 209–210 (1983).

Molecular genetic evidence for heterogeneity in manic depression

Stephen Hodgkinson*†¶, Robin Sherrington*¶, Hugh Gurling*¶ #, Roger Marchbanks†, Stephen Readers‡, Jacques Mallet§, Melvin McInnis||, Hannes Petursson|| & Jon Brynjolfsson||

* Genetics Section, Department of Psychiatry and † Department of Biochemistry, Institute of Psychiatry (University of London), De Crespigny Park, London SE5 8AF, UK

‡ Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

§ Laboratoire de Neurobiologie Cellulaire et Moléculaire, Centre National de Recherche Scientifique, 91190 Gif sur Yvette, France

|| Department of Psychiatry, Borgaspaalinn, Reykjavik, Iceland

Manic depression is a severe cyclic mental illness^{1,2} that can be unipolar or bipolar and has a lifetime risk of approximately 7 per 1,000 in most populations³. Families with multiple cases of manic depression have been described that are compatible with both autosomal dominant and X-linked modes of genetic transmission^{2,4-6}. Psychoactive antidepressant and stimulant drugs that help to ameliorate depression and mania are thought to act by affecting catecholamine neurotransmitter systems such as adrenaline, noradrenaline and dopamine, amongst others⁷. Mutations affecting the tyrosine hydroxylase (*TH*) gene^{8,9}, which encodes the rate-limiting enzyme for the synthesis of these three neurotransmitters⁷, might therefore be responsible for causing the manic depressive phenotype. We have studied three Icelandic kindreds amongst whom it appears that a single autosomal dominant disease allele is segregating. In these families there were 44 cases amongst 73 individuals at risk. Genetic linkage studies were carried out using clones encoding tyrosine hydroxylase^{8,9} the variable portion of the *Harvey-ras-1* (*HRAS1*)¹⁰ locus and the variable region of the insulin gene (*INS*)¹¹. All three markers are closely linked on chromosome 11 (ref. 12) and were used to observe the segregation of restriction fragment length polymorphisms (RFLPs) in the three affected kindreds. We found no evidence for linkage to these markers in any of the three families. In contrast, Gerhard *et al.*¹³⁻¹⁵ found linkage between manic depression and *HRAS1* in a single large Amish kindred. We conclude that there is genetic heterogeneity of linkage in manic depression. Therefore mutations at different loci are responsible for the manic depressive phenotype in the Amish and in Iceland.

The discovery of polymorphic DNA markers has made it possible to begin a systematic search using linkage methods for mutations causing manic depression. However, complicating factors such as heterogeneity of linkage, and incomplete and age-related penetrance need to be overcome. We therefore identified large multiply affected families for whom it was very likely that a single genetic mutation had been responsible for the manic depressive phenotype.

An important issue for psychiatric genetic investigations is the definition of the phenotype to be studied. We employed the widely used Research Diagnostic Criteria (RDC)¹⁶ to establish psychiatric diagnoses. Information required to reach diagnoses was obtained by a structured interview carried out by a psychiatrist with the Lifetime Version of the Schizophrenia and Affective Disorders Schedule (SADS-L)¹⁶ as well as by review of case notes. Summaries of all the clinical histories were compiled and the SADS-L RDC diagnoses were confirmed by two additional psychiatrists. Three Icelandic kindreds (F2, F6, F13) were identified and extensive psychiatric information on past generations was obtained in order to establish firmly that there

Table 1 Lod scores between *HRAS1* and manic depression at various values of recombination fraction (θ)

	Recombination fraction (θ)						
	0.00	0.001	0.05	0.10	0.20	0.30	0.40
Penetrance							
Model I	— ∞	-18.01	-5.10	-3.07	-1.16	-0.28	+0.06
Model II	— ∞	-16.78	-4.99	-2.99	-1.11	-0.24	+0.08
Model III	— ∞	-10.02	-3.24	-1.98	-0.72	-0.11	+0.10
Model IV	— ∞	-12.31	-4.41	-2.72	-1.04	-1.04	+0.06

In two of the three families, six of the cases were unipolar (UP) depressives who were also alcoholic. Because there is a complex interaction between alcoholism and depression²⁵, we excluded, on an *a priori* basis, all cases of alcoholism for the purposes of linkage analysis. After these exclusions, there were 38 cases out of 62 individuals at risk, of these 20 were bipolar (BP) cases and 18 were UP. This proportion is not significantly different from that expected under an autosomal dominant model. Allele frequencies for *HRAS1* were based on data from White *et al.*¹¹. Gene frequency for manic depression in the Icelandic population was estimated to be 0.01²⁶. To test the effect of various penetrance functions on the outcome of the study, two-point lod scores were computed using four different penetrance models. In each model penetrance was assumed to be a function of age and an age-of-onset curve was constructed. In model I, the age-of-onset curve was obtained for each family individually by dividing the total number of affected cases by the number who were affected before the ages of 19, 29, 39, 49, 59 and 79 years. Under this assumption penetrance values for the six age groups were, for pedigree F2, 0.43, 0.86, 0.86, 1.00, 1.00, 1.00; for pedigree F6, 0.06, 0.45, 0.8, 1.00, 1.00, 1.00; for pedigree F13: 0.16, 0.76, 0.92, 1.00, 1.00, 1.00. In model II, penetrances were calculated for all three families combined: 0.14, 0.51, 0.70, 0.89, 1.00, 1.00. In model III, penetrances were set equal to those found in the Amish study: for ages 15-19, 0.10; 20-24, 0.27; 25-29, 0.56, and 30+, 0.63. Two-point lod scores were also calculated for complete penetrance (model IV).

was a single unilateral source of manic depression. One of the pedigrees (F2) is shown in Fig. 1.

Genomic DNA was digested with *HinfI*, *BglII* and *PvuII* restriction enzymes (Amersham) and separated by electrophoresis on agarose gels¹⁷. DNA was transferred by Southern blotting¹⁸ to nylon filters (Hybond N). Pedigree structure was confirmed by hybridization with mini-satellite^{19,20} probe 33.15. Four chromosome 11 linkage markers were investigated in the families. Erythrocyte MER2 antigen, encoded by a gene closely linked to *TH* on chromosome 11²¹, was not found to be informative. A 1.5-kilobase (kb) fragment (Ty7) of a *TH* cDNA clone was used to detect a *BglII* dimorphism at the *TH* locus¹². The

Table 2 Lod scores for linkage between three marker loci and manic depression for various values of recombination fraction (θ), using model II

	Recombination fraction (θ)						
	0.00	0.001	0.05	0.10	0.20	0.30	0.40
<i>HRAS1</i>							
F2	— ∞	-11.39	-3.05	-1.71	-0.56	-0.09	+0.07
F6	— ∞	-2.36	-0.66	-0.36	-0.09	+0.03	+0.05
F13	— ∞	-3.03	-1.28	-0.92	-0.46	-0.18	-0.04
<i>INS</i>							
F2	—	—	—	—	—	—	—*
F6	— ∞	-9.85	-1.68	-0.54	+0.16	+0.18	+0.03
F13	— ∞	-11.09	-2.79	-1.50	-0.45	-0.07	+0.01
<i>TH</i>							
F2	-0.05	0.05	0.04	0.00	0.08	0.12	0.09
F6	— ∞	-8.39	-1.81	-0.86	-0.21	+0.02	+0.01
F13	— ∞	-4.80	-1.44	-0.89	-0.39	-0.15	-0.03

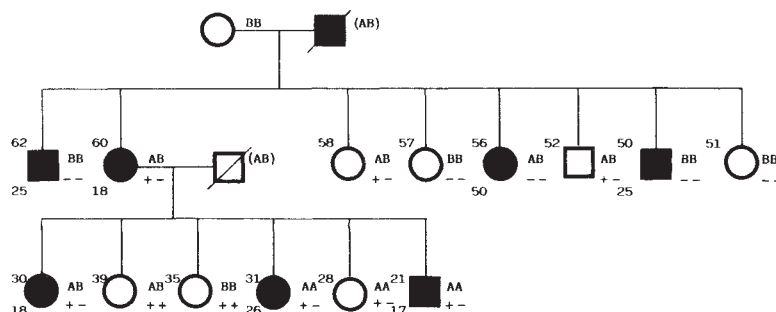
Scores for each family are given separately. An age-of-onset curve was used on the combined penetrance estimates for all 3 families (Model II, as described in Table 1).

* Data not available for F2.

¶ Present address: Molecular Psychiatry Laboratory, Department of Psychiatry, 321 Windeyer Building, Middlesex Hospital Medical School, Mortimer Street, London W1N 8AA, UK.

To whom correspondence should be addressed.

Fig. 1 F2 pedigree structure showing age (above) and age of onset (below) for each individual, *HRAS1* (A, B) and *TH* (+, -) alleles. Squares males; circles, females. Filled symbol, bipolar case; half-filled symbol, unipolar case; open symbol, unaffected; diagonal line, deceased.



variable tandem repeat (VTR) of *HRAS1*¹⁰ was examined in *HinfI*-digested genomic DNA using a 1.4-kb *AvaII* fragment probe containing the VTR 3' to the *HRAS1* locus and four alleles were observed. In two families, F6 and F13, an 879-base-pair (bp) *PvuII* fragment containing the 5' hypervariable region flanking the insulin (*INS*) gene^{11,22} was used to detect 5 different alleles in *PvuII*-digested genomic DNA. The *HRAS1* and *TH* genotypes in F2 are shown in Fig. 1.

The computer package LINKAGE²³ was used to calculate lod scores between disease and marker loci (lod scores quantify the likelihood of linkage between these loci) under different penetrance models. Table 1 gives lod scores for linkage between *HRAS1* and manic depression summed for all families for each model of penetrance. These data exclude the possibility of close linkage between manic depression, *TH*, *INS* and *HRAS1* provided that there is no genetic heterogeneity between families. To assess this possibility, two-point lod scores for the three markers were calculated for each family separately (Table 2). These data provide no evidence for linkage between the disease locus and any of the markers in any of the three families studied. The effect of the exclusion of alcoholics on the linkage analysis was assessed by computing lod scores with all members of the pedigree included. No evidence for linkage was found (data not shown). The three marker loci are closely linked^{11,12}, so multi-point analysis was carried out using the program LINKMAP so the data contributed by each locus could be combined. Linkage between manic depression and the marker loci was excluded ($Z < -2.00$) for a region of ~20 centimorgans on both sides of *HRAS1*.

We conclude that mutations at different loci are causing manic depression in the Old Order Amish and in Iceland. Synthetic gene frequency maps show that the Icelandic population shares the greatest genetic similarity with the populations of northern Germany and with other northern Europeans²⁴; the Amish are known to have migrated from Germany. Therefore further research is required to establish the degree of genetic heterogeneity of linkage that is involved in manic depression.

We are grateful to Icelandair for kindly transporting samples to London, to Dr P. Tippet (MRC Blood Group Unit) for MER2 genotyping, to Andrew King for help with computing and to Dr Robin Winter for advice. H.M.D.G. is a Wellcome senior fellow in clinical science. The research was supported by the Wellcome Trust and the MRC of the UK.

- Mallet, J. *et al.* in *Role of RNA and DNA in Brain Function; a Molecular Biological Approach* (ed. Guiditta, A. *et al.*) 57-70 (Nijhoff, Norwell, Massachusetts, 1987).
- Krontiris, T. G., DiMartino, N. A., Colb, M. & Parkinson, D. R. *Nature* **313**, 369-374 (1985).
- Bell, G. I., Selby, M. J. & Rutter, W. J. *Nature* **295**, 31-35 (1982).
- Moss, P. A. H., Davies, K. E., Bon, I. C., Mallet, J. & Reeders, S. T. *Nucleic Acids Res.* **14**, 9927-9932 (1986).
- Gerhard, D. S. *et al.* *Am. J. hum. Genet.* **36**, S35 (1984).
- Gerhard, D. S. *et al.* *Symp. Genet. Res. Psychiat.* Berlin (1986).
- Kidd, K. K., Gerhard, D. S., Kidd, J. R., Housman, D. & Egeland, J. *Clin. Neuropharmacol.* **7**, S1, 198-199 (1984).
- Spitzer, J., Endicott, J. & Robins, E. *Archs gen. Psychiat.* **35**, 773-782 (1978).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor, New York, 1982).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J. & Ponder, B. A. J. *Am. J. hum. Genet.* **39**, 11-24 (1986).
- Hodgkinson, S., Gurling, H. M. D., Marchbanks, R. M., McInnis, M. & Petursson, H. *J. psychiat. Res.* (in the press).
- Palmer, D. K. *et al.* *Cytogenet. Cell Genet.* **40**, 370 (1985).
- Rotwein, P., Yokoyama, S., Didier, D. K. & Chirgwin, J. M. *Am. J. hum. Genet.* **39**, 291-299 (1986).
- Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3443-3446 (1984).
- Piazza, A., Menozzi, P. & Cavalli-Sforza, L. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2638-2642 (1981).
- Gurling, H. M. D., Oppenheim, B. E. & Murray, R. M. *Acta Genet. med. Gemell.* **33**, 333-339 (1984).
- Helgason, T. *Acta psychiatr. scand. Suppl.* **162**, 81-90 (1961).

Close linkage of c-Harvey-ras-1 and the insulin gene to affective disorder is ruled out in three North American pedigrees

Sevilla D. Detera-Wadleigh, Wade H. Berrettini, Lynn R. Goldin, Denise Boorman, Stewart Anderson* & Elliot S. Gershon

Clinical Neurogenetics Branch, Division of Intramural Research Programs, National Institute of Mental Health, 9000 Rockville Pike, Bethesda, Maryland 20892, USA

* University of Connecticut Health Center, School of Medicine, Farmington, Connecticut 06032, USA

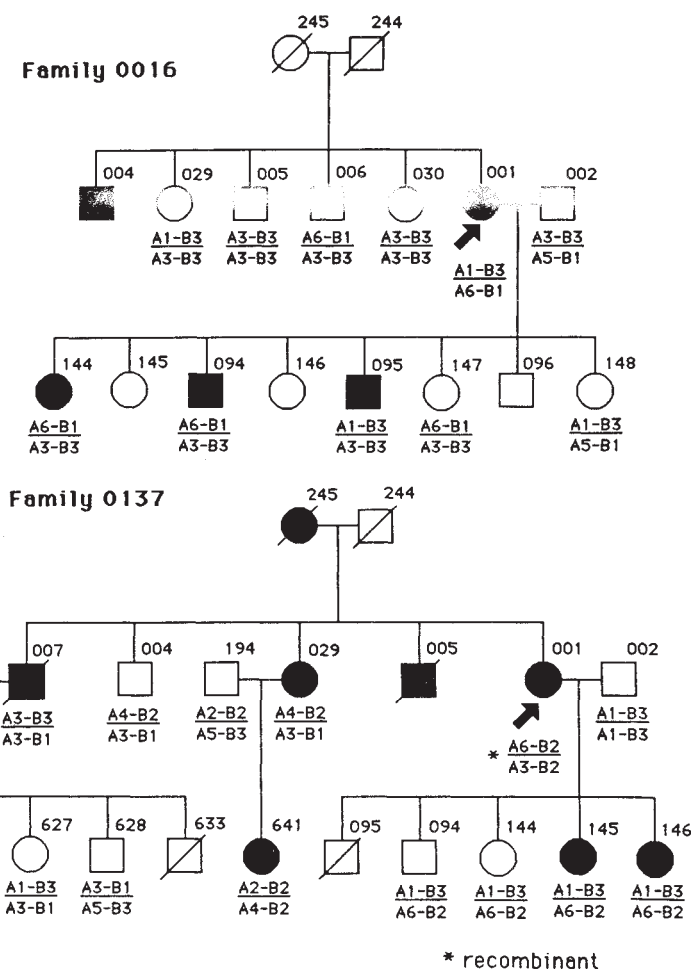
Affective disorder (AD) is one of the major forms of functional psychoses. Although the mode of transmission is uncertain, family, twin and adoption studies strongly suggest a genetic involvement (see ref. 1 for review). Because a basic biochemical abnormality is not known, direct analysis of the disease using a probe for the defective gene is not possible. However, a specific locus can be tested for its relevance to the aetiology of AD by genetic linkage, using restriction fragment length polymorphisms (RFLPs). Using probes for the c-Ha-ras-1 oncogene and the insulin gene, Gerhard *et al.*² and Egeland *et al.*³ found convincing evidence for close linkage between these markers and a locus for AD in a large Old Order Amish pedigree. In an attempt to confirm this finding, we examined three bipolar pedigrees outside the Amish population. Our results indicate the absence of linkage from 0 to 15% recombination frequency between AD and the insulin gene-*HRAS1* region in these pedigrees.

Pedigrees in which several members have major affective disorders were recruited through bipolar patients at the National

Received 4 December 1986; accepted 23 January 1987.

- Perris, C. in *Genetic Aspects of Affective Illness* (ed. Perris, C.) 9-19 (SP Medical and Scientific, New York, 1979).
- Nurnberger, J. I. & Gershon, E. S. in *Handbook of Affective Disorders* (ed. Paykel, E. S.) 126-145 (Churchill Livingstone, London, 1982).
- Boyd, J. H. & Weissman, M. W. *Archs gen. Psychiat.* **38**, 1039-1046 (1981).
- Mendelwicz, J. *IVth Wld Congr. Biol. Psychiat.* Philadelphia 306 (1985).
- Winokur, G. & Clayton, P. in *Recent Adv. Biol. Psychiat.* **9**, 35-50 (Plenum, New York, 1967).
- Winokur, G., Clayton, P. & Reich, T. in *Manic Depressive Illness* (eds Winokur, G., Clayton, P. & Reich, T.) 1-150 (Mosley, St Louis, 1969).
- Cooper, J. R., Bloom, F. E. & Roth, R. H. in *The Biochemical Basis of Neuropsychopharmacology* (eds Cooper, J. R., Bloom, F. E. & Roth, R. H.) 1-367 (Oxford University Press, New York, 1982).
- Lamoureux, A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **79**, 3881-3885 (1982).

Fig. 1 Segregation of the insulin gene and *HRAS1* haplotypes in two pedigrees with major affective disorder. A1-6 and B1-3 are the haplotypes of the insulin gene and *HRAS1* (see text). Affected members are denoted by solid symbols; circles, females; squares, males; diagonal line, deceased; arrow, propositions. Genomic DNAs from indicated individuals were prepared either directly from peripheral blood or from cultured lymphoblasts transformed by the Coriell Institute for Medical Research, Human Genetic Mutant Cell Repository, Camden, New Jersey, USA. One set of aliquots of DNA was digested with *PvuII* and another with *SacI*. The DNA fragments were electrophoresed on agarose and transferred to nylon membranes (ICN) by the method of Southern²⁰. The *PvuII* blots were hybridized with the ³²P-labelled insert from phins 310 (a generous gift of Dr Graeme Bell, Chiron Corporation) and the *SacI* blots were hybridized with ³²P-labelled EJ6.6¹² insert (American Type Culture Collection). The DNAs were labelled according to Feinberg and Vogelstein^{21,22} to a specific activity of 10^9 c.p.m. μg^{-1} . Hybridization was performed for 48 h at 42 °C with 50% formamide, $5 \times \text{SSC}$, $2 \times \text{Denhardt's}$, 20 mM Na^+ phosphate, pH 6.5, $200 \mu\text{g ml}^{-1}$ denatured salmon sperm DNA, 0.1% SDS and 10% dextran sulphate and denatured ³²P-labelled probe. The filters were washed at room temperature with $2 \times \text{SSC}$, 0.1% SDS three times, 15 min each time, followed by at least one wash at 65 °C for 1 h. Autoradiography was done using Kodak XAR-5 films with Dupont Lightning Plus intensifying screens at -70 °C for at least 2 days.



Institute of Mental Health. Diagnoses were determined using modified Research Diagnostic Criteria based on the Schedule for Affective Disorders and Schizophrenia Lifetime version⁴ interview, information from relatives, and review of all available hospital records⁵. In the present study individuals with a diagnosis of bipolar I (BPI) are classified as 'affected'. Also included in this category are bipolar II (BPII) with major depression and unipolar (UP), based on family and twin studies which find these diagnoses greatly increased in relatives of manic depressives as compared with relatives of normal controls^{1,5}. Schizoaffective and anorexic relatives are classified as affected because of the high prevalence of BP and UP disorders in families of these patients^{5,6}. Family members with minor depression alone were classified as 'unaffected'. The pedigrees collected were North American caucasians of European origin.

Attempts to correlate specific loci with AD have yielded primarily negative results^{7,8}. The *HLA* region was initially believed to be loosely linked to AD but later studies by other investigators failed to confirm this finding (see ref. 7 for review). Other studies have suggested X-linkage in a proportion of families with AD, but the data have not been consistently reproducible⁷. We have obtained evidence suggesting that somatostatin and neuropeptide Y RFLPs are not linked to an AD locus (unpublished results). Proopiomelanocortin RFLPs were found not to be associated with bipolar illness⁸.

Recent data obtained from the study of an Amish pedigree demonstrate linkage of BP and UP affective disorders to the insulin gene-*HRAS1* region of chromosome 11³. The multiple alleles of the insulin and *HRAS1* genes⁹⁻¹¹ make this region of chromosome 11 extremely useful for genetic linkage study. The high degree of polymorphism at the 5' flanking region of the insulin gene results from the variation in the number and arrangement of tandemly repeated nucleotides¹⁰. When *PvuII*

is used, up to six different restriction fragments can be detected using phins 310 as probe and these alleles are designated A1 to A6⁹⁻¹¹.

The oncogene *HRAS1*, in close proximity to the insulin gene, has been shown to be highly polymorphic (ref. 9 and D. S. Gerhard, personal communication). Multiple polymorphic fragments, designated here as B alleles, have been observed when *SacI* is used as the enzyme and EJ6.6 as probe (ref. 12 and D. S. Gerhard, personal communication).

Two of the three pedigrees examined for insulin and *HRAS1* RFLPs are shown in Fig. 1. The total number of individuals examined in all three pedigrees was 38, 15 of whom were affected. As indicated, haplotypes were deduced from the genotypes obtained with the insulin gene and *HRAS1*. This was done because the genetic distance between these genes is only 3%¹³. In the three pedigrees studied only one recombinant was found, individual 137001 (Fig. 1). She was assigned a haplotype according to her *HRAS1* genotype because the results from the Amish study^{2,3} indicate close linkage of AD to the *HRAS1* gene. The corresponding age and diagnosis of each family member studied are summarized in Table 1.

Linkage tests between marker haplotypes and affective illness were performed by estimation of lod scores using the LIPED¹⁴ program. A correction for age was incorporated in the calculations. For example, the normal diagnosis for individuals 16147 and 16148 could change with time as they are just approaching the age of risk. Population prevalence also enters into lod score computations, in assigning a gene frequency for the disease allele. For BP illness and acute schizoaffective disorder, the combined population prevalence is ~1%⁵. For UP illness, varying rates have been reported which are a function of the diagnostic criteria used, birth cohort and other factors^{15,16}. With stringent criteria, we have estimated the population rate to be 5.8%⁵.

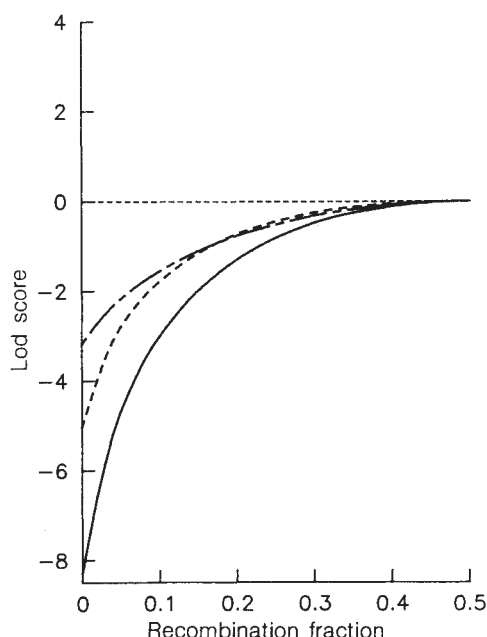


Fig. 2 Graph of the combined lod scores from 3 pedigrees as a function of recombination fraction to determine linkage between an AD locus and the insulin gene-*HRAS1* region. The lod scores were calculated based on haplotypes determined from insulin and *HRAS1* genotypes as shown in Fig. 1. The computer program LIPED^{14,23} was used under three models with a straight-line age correction (age of risk specified as 12 to 65 years). These models were: (1) dominant with reduced penetrance (long and short dashed line), where $q_a = 0.035$, $f_{aa} = f_{AA} = 0.5$, $f_{AA} = 0.005$; (2) recessive (short dashed line), where $q_a = 0.20$, $f_{aa} = 0.9$, $f_{AA} = f_{AA} = 0.005$ and; (3) dominant model (solid line) using assumptions from the Amish pedigree where strong evidence for linkage was found^{2,3}, that is, $q_a = 0.015$, $f_{aa} = f_{AA} = 0.85$, $f_{AA} = 0.001$, age of risk specified as 15 to 45 years. Genotype penetrances are denoted by f_{aa} for abnormal homozygote, f_{AA} for the heterozygote and f_{AA} for the normal homozygote: q_a is the frequency of the recessive allele. Linkage is rejected at the recombination fractions with lod scores less than -2.0 (ref. 19). Close linkage is rejected for all models tested.

Similar rates have been obtained in other independent studies^{17,18}. Up to four times higher rates have been reported using more inclusive criteria¹⁵. For the linkage analyses presented here, we assume that these 'high-density' families represent a rarer subtype of AD, and have based our genetic models on a prevalence rate of 3.5%, which is half our estimated combined prevalence rate for BP, schizoaffective and UP disorders. However, the lod scores are not substantially changed within a considerable range of gene frequencies (data not shown).

To compare our results with those obtained by Gerhard *et al.*² and Egeland *et al.*³, a dominant mode of inheritance with 85% penetrance was assumed. At 0–15% recombination the combined lod scores obtained from the three pedigrees examined in the present study were less than -2 , which allows rejection of close linkage¹⁹. The results are graphically presented in Fig. 2. A similar observation was seen when pedigree 137 was examined alone (data not shown). Negative lod scores were also obtained when two other models, dominant with reduced penetrance and recessive, were tested (Fig. 2). When the insulin and *HRAS1* loci were analysed separately in these pedigrees, assuming in each case the mode of inheritance used in the Amish study^{2,3}, linkage to AD is rejected at less than 5% recombination (data not shown). When the haplotype of individual 137001 was assigned according to her insulin genotype the change in the lod score values was not significant (data not shown). If we narrow our classification of affective diagnoses to include only BP, UP and schizoaffective or only BP and schizoaffective the

Table 1 Diagnoses and ages of pedigree members

Pedigree 16		
Individual no.	Age	Diagnosis
001	54	BPI
002	52	unaffected
005	59	unaffected
006	57	unaffected (minor depression)
029	61	unaffected
030	51	unaffected
094	28	BPI, drug abuse
095	24	BPI
144	32	BPI
147	23	unaffected
148	21	unaffected

Pedigree 137		
Individual no.	Age	Diagnosis
001	65	BPII
002	64	unaffected
004	73	unaffected
007	68	BPI
029	70	BPI
094	38	unaffected
144	34	unaffected
145	28	BPII, major depression, anti-social, bulimia, drug abuse, alcoholism
146	26	minor depression, anorexia, bulimia
194	65	unaffected (alcoholism)
219	65	unaffected (minor depression)
626	30	probable UP
627	33	unaffected
628	34	unaffected (panic disorder)
641	32	schizoaffective, manic, depressed

The family members included here are only those for which DNAs were available. The ages shown were either at last examination or death.

lod scores are slightly higher but our conclusions are not changed (data not shown). Our data, therefore, support the hypothesis of independent segregation of the insulin gene-*HRAS1* region and affective illness. This contrasts with the evidence reported earlier^{2,3}. The reason for this discrepancy is not clear. Conceivably these contrasting findings favour aetiological heterogeneity in AD. The AD in the Amish pedigree may represent a genetically distinct form of the illness. However, the hypothesis of genetic heterogeneity will remain speculative until additional pedigrees and more regions of the genome are studied.

Received 11 December 1986; accepted 26 January 1987.

- Nurnberger, J. I., Jr, Goldin, L. R. & Gershon, E. S. in *The Medical Basis of Psychiatry* (eds Winokur, G. & Clayton, P.) 486–521 (Saunders, Philadelphia, 1986).
- Gerhard, D. S. *et al. Am. J. hum. Genet.* **36**, 3S (1984).
- Egeland, J. A. *et al. Nature* **325**, 783–787 (1987).
- Spitzer, R. & Endicott, J. *Critical Issues in Psychiatry Diagnosis* (Raven Press, New York, 1978).
- Gershon, E. S. *et al. Archs gen. Psychiat.* **39**, 1157–1167 (1982).
- Gershon, E. S. *et al. Am. J. Psychiat.* **141**, 1419–1422 (1984).
- Goldin, L. R. & Gershon, E. S. *Psychiat. Developments* **1**, 387–418 (1983).
- Feder, J., Gurling, H. M. D., Darby, J. & Cavalli-Sforza, L. L. *Am. J. hum. Genet.* **37**, 286–294 (1985).
- Willard, H. F., Skolnick, M. H., Pearson, P. L. & Mandel, J. L. *Cytogenet. Cell Genet.* **40**, 360–489 (1985).
- Bell, G. I., Selby, M. J. & Rutter, W. J. *Nature* **295**, 31–35 (1982).
- Bell, G. I., Karam, J. H. & Rutter, W. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5759–5763 (1981).
- Shih, C. & Weinberg, R. A. *Cell* **29**, 161–169 (1982).
- Grzeschik, K. H. & Kazazian, H. H. *Cytogenet. Cell Genet.* **40**, 179–205 (1985).
- Ott, J. *Am. J. hum. Genet.* **26**, 588–597 (1974).
- Gershon, E. S. *Archs gen. Psychiat.* **41**, 104–105 (1985).
- Klerman, G. L. *et al. Archs gen. Psychiat.* **42**, 696–702 (1985).
- Tsuang, M. T., Winokur, G. & Crowe, R. R. *Br. J. Psychiat.* **137**, 497–504 (1980).
- Weissman, M. M. *et al. Archs gen. Psychiat.* **41**, 13–21 (1984).
- Morton, N. E., *Am. J. hum. Genet.* **7**, 277–318 (1955).
- Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **137**, 266–267 (1984).
- Hodge, S. E., Morton, L. A., Tideman, S., Kidd, K. K. & Spence, M. A. *Am. J. hum. Genet.* **31**, 761–762 (1979).

Cholecystikinin induces a decrease in Ca^{2+} current in snail neurons that appears to be mediated by protein kinase C

Constance Hammond, Danièle Paupardin-Tritsch, Angus C. Nairn* Paul Greengard* & Hersch M. Gerschenfeld

Laboratoire de Neurobiologie, Ecole Normale Supérieure, 46 Rue d'Ulm, 75230 Paris Cedex 05, France

*Laboratory of Molecular and Cellular Neuroscience, The Rockefeller University, 1230 York Avenue, New York, New York 10021-6399, USA

Three distinct classes of protein kinases have been shown to regulate Ca^{2+} current in excitable tissues. Cyclic AMP-dependent protein kinase mediates the action of noradrenaline on the Ca^{2+} current of cardiac muscle cells^{1,2}. Cyclic GMP-dependent protein kinase mediates the serotonin-induced modulation of the Ca^{2+} current in identified snail neurons³. The Ca^{2+} /diacylglycerol-dependent protein kinase (protein kinase C) has also been found to regulate Ca^{2+} currents of neurons^{4,5}. However, no neurotransmitter has yet been shown to regulate Ca^{2+} current through the activation of protein kinase C. We now report that cholecystikinin, a widely occurring neuropeptide⁶ which is present in molluscan neurons^{7,8}, modulates the Ca^{2+} current in identified neurons⁹ of the snail *Helix aspersa*, and that this effect appears to be mediated by protein kinase C. Specifically, sulphated cholecystikinin octapeptide 26–33 (CCK8), activators of protein kinase C, and intracellular injection of protein kinase C, all shorten the Ca^{2+} -dependent action potential and decrease the amplitude of the Ca^{2+} current in these cells. All these effects are not reversible within the duration of the experiments. Moreover, intracellular injections of low concentrations of protein kinase C, which are ineffective by themselves, enhance the effectiveness of low concentrations of CCK8 on the Ca^{2+} current.

The effect of CCK8 on the somatic Ca^{2+} -dependent action potential was measured in the identified snail neurons F77 (Fig. 1a) and D2 (not shown) bathed in tetraethylammonium (TEA)/ Cs^+ -saline. CCK8 (5 nM) reduced the duration of the Ca^{2+} -dependent action potential, a maximum decrease being measured 7 min after bath application of the peptide. The effect of CCK8 was not reversed by washing the preparation for 45 min (not shown).

We next analysed the effect of CCK8 on the ionic currents in these cells. The effect of CCK8 on the inward current was measured in F77 and D2 neurons bathed in tetrodotoxin (TTX)/TEA/ Cs^+ / Ba^{2+} -saline. CCK8 (5 nM) decreased the inward current amplitude, a stable 20–25% decrease being obtained 7 min after the peptide application (Fig. 1b). No recovery occurred even after washout of the peptide for 40 min. The threshold concentration of CCK8 for an effect on the inward current was 0.5 nM, and a maximal effect (reduction of 70–80%) was obtained after 1–2 min with 500 nM CCK8. The $I-V$ curves shown in Fig. 1c confirm that CCK8 decreases the inward current through Ca^{2+} -channels. CCK8 (5 nM and 20 nM) decreased the inward current at all membrane potential levels examined. The membrane potential level at which the inward current was maximal was the same in control and treated neurons, and at this potential value, CCK8 induced a maximal decrease in inward current. CCK8 action on the inward current through Ca^{2+} -channels was confirmed using the ionophore nystatin to exchange intracellular K^+ for Cs^+ (see ref. 10), and the extracellular Ca^{2+} was replaced by Ba^{2+} and NaCl by TrisCl. Under these conditions, in which Cs^+ was the only intracellular monovalent cation and K^+ and Na^+ were absent from the extracellular medium, the effect of CCK8 on the amplitude of

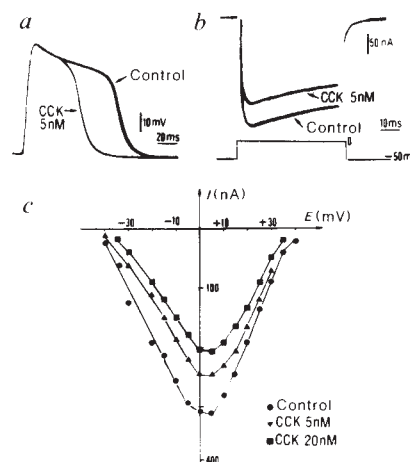


Fig. 1 Effect of CCK8 on the Ca^{2+} -dependent action potential and the inward current recorded in identified snail neurons. **a**, A control action potential showing a large Ca^{2+} -dependent plateau, was recorded from a F77 neuron bathed in TEA/ Cs^+ -saline. Exposure to 5 nM CCK8 elicited a decrease in the duration of the action potential. Scale bars are 10 mV (vertical) and 20 ms (horizontal). **b**, A control inward current was recorded from a different F77 neuron bathed in TTX/TEA/ Cs^+ / Ba^{2+} -saline. Exposure to 5 nM CCK8 elicited a decrease in the peak amplitude of the inward current. A constant value was reached after 7 min. Scale bars are 50 nA and 10 ms. **c**, $I-V$ curves relating the inward current peak amplitudes and the membrane potential at which they were recorded in a D2 neuron bathed in TTX/TEA/ Cs^+ / Ba^{2+} -saline. The values of the peak inward currents were obtained in the absence (circles) or presence of 5 nM (triangles) and 20 nM (squares) CCK8. The three curves were corrected for linear leakage conductance.

Methods. The isolated and desheathed abdomino-visceral ganglionic mass of the snail *Helix aspersa* was bathed in saline of the following composition (mM): NaCl 120, KCl 2.5, CaCl_2 6, MgCl_2 3.5, HEPES 10 (pH 7.4). Identified snail neurons (D2 and F77, ref. 9) were impaled with two microelectrodes connected to a current/voltage clamp system. To unmask the Ca^{2+} -dependent component of the action potential, some of the K^+ -conductances were blocked by replacing one-fourth of the NaCl with 30 mM TEA Cl and replacing KCl by CsCl in the saline (TEA/ Cs^+ -saline). To isolate the inward current through Ca^{2+} -channels, the neurons were bathed in TEA/ Cs^+ saline which contained 1 μM TTX, and in which 6 mM Ca^{2+} was replaced by 6 mM Ba^{2+} (TTX/TEA/ Cs^+ / Ba^{2+} -saline). To record the inward current in **b** and Figs 2–4, the neurons were held at -50 mV and depolarized to 0 mV for 60 ms. CCK8 sulphate (1 mM stock solution in 1 M Tris-Cl, pH 7.8 was stored at -20°C and a new sample used in each experiment.

the inward current was unchanged. We also tested the effect of the peptide on the TEA-insensitive, Ca^{2+} -independent outward current in F77 and D2 neurons bathed in TEA/ Cs^+ -saline in which Ca^{2+} was replaced by Mg^{2+} . In these experiments ($n = 4$), the cells were held at -50 mV and depolarized to $+20$ mV for 1 s. The application of CCK8 (5 nM) had no effect on the outward current in any of these experiments (not shown).

The intracellular mechanisms which might mediate the effect of CCK8 were further investigated by making prolonged repetitive injections of either cAMP or cGMP, but neither affected the inward Ca^{2+} current in either D2 or F77 neurons. Intracellular injection of EGTA, a Ca^{2+} chelator (Fig. 2a), both before and during the application of CCK8 (5 nM) either totally prevented or markedly attenuated the inward current decrease induced by the peptide. If the injection of EGTA was stopped either before or during washout of the peptide, a delayed but full effect of CCK8 was observed (not shown). Moreover, injection of EGTA had no effect when CCK8 had previously decreased the inward current. The efficacy of the EGTA injections was tested in parallel experiments in which intracellular

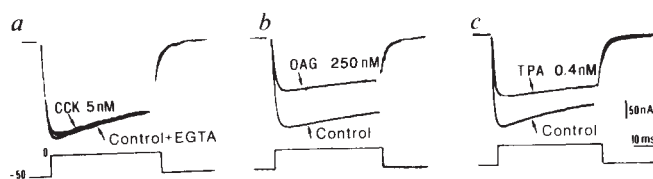


Fig. 2 The effect of CCK8 on the inward Ca^{2+} current of identified snail neurons is blocked by EGTA and is mimicked by activators of protein kinase C. *a*, A control inward current was recorded from a F77 neuron bathed in TTX/TEA/ Cs^+ / Ba^{2+} -saline. EGTA was pressure injected (pulses of 2 bars for 50 ms at 0.5 Hz) for 15 min. Five minutes after the onset of the EGTA injection, CCK8 (5 nM) was applied. The trace corresponds to two superimposed recordings obtained before and 7 min after CCK8 application. *b*, A control inward current was recorded from a F77 neuron bathed in TTX/TEA/ Cs^+ / Ba^{2+} -saline. Bath application of OAG (250 nM) markedly reduced the amplitude of the inward current. *c*, A control inward current was recorded from a different F77 neuron bathed in TTX/TEA/ Cs^+ / Ba^{2+} -saline. Bath application of phorbol ester TPA (0.4 nM) reduced the amplitude of the inward current. EGTA was pressure injected from a micropipette containing 0.25 M $\text{K}_2\text{-EGTA}$ /60 mM KCl/1 mM HEPES (pH 7.4). OAG (10 mg ml^{-1}) and phorbol ester TPA (1 mg ml^{-1}) were dissolved in absolute ethanol and diluted in TTX/TEA/ Cs^+ / Ba^{2+} -saline. Scale bars 50 nA and 10 ms.

injections of EGTA were shown to block the Ca^{2+} -dependent outward current evoked by repetitive depolarizing pulses in the identified neurons bathed in normal saline (not shown).

The EGTA experiments suggested a participation of cytosolic Ca^{2+} in mediating the effect of CCK8. Because inositol trisphosphate (InsP_3) has been shown to control the release of Ca^{2+} into the cytosol¹¹, we injected InsP_3 into the identified neurons. It had no consistent effects either on the basal inward current or on the CCK8-induced decrease of inward current. Cytosolic Ca^{2+} is also involved in the activation of protein kinase C¹², so possible involvement of this enzyme was examined. Bath application of oleoylacylglycerol (OAG, 250 nM), a diacylglycerol analogue which activates protein kinase C¹², induced a marked and irreversible decrease in the amplitude of the inward current in F77 neurons (Fig. 2*b*). When the OAG concentration was increased to 60 μM its effect was maximal, the amplitude of the inward current being reduced by 80%. In these conditions, addition of CCK8 to the bath was totally ineffective. In other experiments, application of the phorbol ester 12-*O*-tetradecanoyl-phorbol-13-acetate (TPA, 0.4 nM), another potent activator of protein kinase C¹², also induced a marked and irreversible decrease in the amplitude of the inward current (Fig. 2*c*).

The effect of intracellular pressure injection of protein kinase C was investigated. In all neurons tested ($n = 6$) pressure injection of protein kinase C evoked a marked shortening of the Ca^{2+} -dependent somatic action potential (Fig. 3*a*). Pressure injection of the enzyme also evoked a marked and irreversible decrease of the inward current in all neurons tested ($n = 10$) (Fig. 3*b*). In contrast, repeated pressure injection of heat-inactivated protein kinase C did not affect the amplitude of the inward current in the identified neurons ($n = 7$) (Fig. 3*c*).

Intracellular injection of protein kinase C also enhanced the effectiveness of low concentrations of CCK8. In control F77 neurons ($n = 5$), CCK8 (1 nM) evoked a $21.6 \pm 1.6\%$ (mean $\pm$ s.d.) decrease in the inward current of the identified neurons (Fig. 4*a*). Pressure injection of a fivefold diluted sample of the protein kinase C preparation had no effect by itself on the Ca^{2+} current (Fig. 4*b*). However, in neurons ($n = 5$) which had been injected with the diluted sample of protein kinase C, CCK8 (1 nM) induced a $35.6 \pm 3.4\%$ (mean $\pm$ s.d.) decrease in the amplitude of the inward current (Fig. 4*b*), a decrease significantly greater ($P < 0.001$, Student's test) than that obtained by applying CCK8 alone.

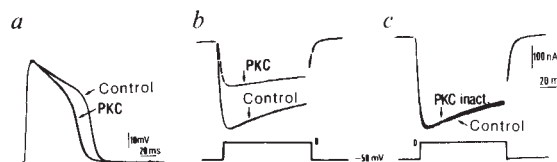


Fig. 3 Intracellular injection of protein kinase C mimics the effect of CCK8 on the Ca^{2+} -dependent action potential and on the inward Ca^{2+} current of identified snail neurons. *a*, A control Ca^{2+} -dependent action potential was recorded from a F77 neuron bathed in TEA/ Cs^+ saline. Protein kinase C was then pressure injected (2 pulses of 2 bar for 10 ms). After 2–3 min, a decrease in the duration of the somatic spike (PKC) was observed. Scale bars 10 mV and 20 ms. *b*, A control inward current was recorded from a different F77 neuron bathed in TTX/TEA/ Cs^+ / Ba^{2+} -saline. Protein kinase C was then pressure injected (3 pulses of 2 bar for 10 ms). After 7 min, a marked decrease of the inward current (PKC) was observed. *c*, Pressure injection (50 pulses of 2 bar for 100 ms) of heat-inactivated protein kinase C in a different F77 neuron had no effect on the inward current. The trace corresponds to two superimposed recordings obtained before and 15 min after (PKC inact.) injection of the enzyme. Scale bars 100 nA and 20 ms. Protein kinase C was purified from bovine brain using a modification of the method of Kikkawa *et al.*²³. The enzyme was totally inactivated by incubation at 60 °C for 12 min. Protein kinase C was pressure injected using siliconized micropipettes of 1–2 μm tip diameter.

We propose the following mechanism to account for the modulation by CCK8 of the inward Ca^{2+} current of D2 and F77 neurons: the peptide binds to receptors on the plasma membrane of the neurons, causing the formation of diacylglycerol from inositol phospholipids; diacylglycerol, in the presence of an adequate cytosolic Ca^{2+} concentration, activates protein kinase C which in turn phosphorylates Ca^{2+} -channel(s) or some protein associated with them. This interpretation is supported by the following findings. (1) CCK8 acts on identified snail neurons at concentrations similar to those at which it binds to receptors in both pancreatic¹³ and brain¹⁴ membranes. (2) Two activators of protein kinase C (OAG and the phorbol ester TPA) mimic the effect of CCK8 on the Ca^{2+} current. (3) Intracellular injection of protein kinase C both mimics and enhances the CCK8-induced decrease of Ca^{2+} current. (4) Intracellular injections of EGTA prevent the CCK8-induced decrease of Ca^{2+} current. This effect of EGTA may be explained by the fact that protein kinase C, even in the presence of diacylglycerol, requires low concentrations of Ca^{2+} for activity¹².

The decrease in inward Ca^{2+} current induced by CCK8 appears to be irreversible or, at least, reversible only very slowly. In the same identified snail neurons, a brief application of dopamine evokes a decrease of the Ca^{2+} current which is rapidly reversible and not affected by the intracellular injection of EGTA¹⁵. These results suggest that two distinct mechanisms

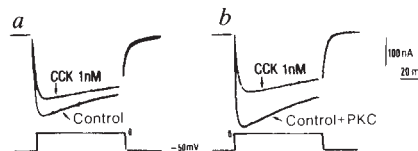


Fig. 4 Injection of protein kinase C enhances the effectiveness of CCK8 on the inward Ca^{2+} current of identified snail neurons. *a*, Inward currents recorded from a F77 neuron bathed in TTX/TEA/ Cs^+ / Ba^{2+} -saline before and 7 min after exposure to 1 nM CCK8; the peptide reduced the inward current by 20%. *b*, Protein kinase C, diluted fivefold, was pressure injected into a different F77 neuron (50 pulses of 2 bar for 20 ms) before bath application of CCK8. This concentration of protein kinase C had no effect on the inward current; the control trace actually corresponds to two superimposed recordings obtained before and 10 min after injection of protein kinase C. CCK8 (1 nM) was then applied; the inward current measured 7 min after the peptide application was decreased by 35%. Scale bars 100 nA, 20 ms.

exist for the regulation of the Ca^{2+} current, one possibly involved in long-term and the other in short-term changes of neuronal excitability.

Activators of protein kinase C have previously been found to increase the Ca^{2+} current in *Aplysia* bag cells^{4,16}, and to decrease reversibly the Ca^{2+} current in cultured chick embryo dorsal root ganglion neurons⁵. However, the neurotransmitters which activate protein kinase C are not known in any of these systems. The opposite actions of protein kinase C in these different systems suggest that distinct types of Ca^{2+} -channels may be phosphorylated by the enzyme. Protein kinase C has also been reported to decrease two different K^{+} -currents in photoreceptors of the mollusc *Hermissenda*¹⁷ and a Cl^{-} -current in mammalian hippocampal neurons¹⁸. More recently, it has been reported¹⁹ that bradykinin depresses the M current, a particular K^{+} -current, in a neuroblastoma hybrid cell line in culture and that this effect is mimicked by activators of protein kinase C. CCK8 has been shown to depolarize and/or excite mammalian hippocampal pyramidal cells²⁰, dentate gyrus²¹ and spinal dorsal neurons²². However, the mechanism of action of CCK8 in these neurons is not known. We have provided evidence in this study suggesting that, at least in identified neurons of *Helix aspersa*, CCK8 modulates the Ca^{2+} current by activating protein kinase C.

This work was supported by grants from the Centre National de la Recherche Scientifique, France (UA 295 and ATP Pharmacologie de la Neurotransmission) and from the Université P. and M. Curie, Paris to H.M.G. and D.P.-T., and from the USPHS

(MH 40899) to A.C.N. and P.G. We thank Mr S. Lucania of Ely Lilly & Sons, and Drs B. Roques and D. Pelaprat of the Faculté de Pharmacie, Paris for gifts of CCK8.

Received 26 November 1986; accepted 2 January 1987.

- Osterrieder, W. *et al.* *Nature* **298**, 576-578 (1982).
- Brum, G., Flockerzi, V., Hofman, F., Osterrieder, W. & Trautwein, W. *Pflügers Arch. ges. Physiol.* **398**, 146-154 (1983).
- Paupardin-Tritsch, D., Hammond, C., Gerschenfeld, H. M., Nairn, A. C. & Greengard, P. *Nature* **323**, 812-814 (1986).
- DeRiemer, S. A., Strong, J. A., Albert, K. A., Greengard, P. & Kaczmarek, L. K. *Nature* **313**, 313-316 (1985).
- Rane, S. G. & Dunlap, K. *Proc. natn. Acad. Sci. U.S.A.* **83**, 184-188 (1986).
- Crawley, J. N. *Ann. N.Y. Acad. Sci.* **448**, 1-8 (1985).
- Osborne, N. N., Cuello, A. C. & Dockray, C. G. *Science* **216**, 409-411 (1982).
- Ono, J. K. *Neuroscience* **18**, 957-974 (1986).
- Kerkut, G. A., Lambert, J. C. D., Gayton, R. J., Loker, J. E. & Walker, R. J. *Comp. Biochem. Physiol.* **50A**, 1-25 (1975).
- Tillotson, D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1497-1500 (1979).
- Berridge, M. J. *Biochem. J.* **220**, 345-360 (1984).
- Nishizuka, Y. *Science* **225**, 1365-1370 (1984).
- Sankaran, H., Goldfine, I. D., Deveney, C. W., Wong, K.-Y. & Williams, J. A. *J. biol. Chem.* **255**, 1849-1853 (1980).
- Williams, J. A., Vigna, S. R., Sakamoto, C. & Goldfine, I. D. *Ann. N.Y. Acad. Sci.* **448**, 220-230 (1985).
- Paupardin-Tritsch, D., Colombaioni, L., Deterre, P. & Gerschenfeld, H. M. *J. Neurosci.* **5**, 2522-2532 (1985).
- Strong, J. A., Fox, A. P., Tsien, R. W. & Kaczmarek, L. K. *Biophys. J.* **49**, 430a (1986).
- Farley, J. & Auerbach, S. *Nature* **319**, 220-223 (1986).
- Madison, D. V., Malenka, R. C. & Nicoll, R. A. *Nature* **321**, 695-697 (1986).
- Higashida, H. & Brown, D. A. *Nature* **323**, 333-335 (1986).
- Dodd, J. & Kelly, J. S. *Brain Res.* **205**, 337-350 (1981).
- Brooks, P. A. & Kelly, J. S. *Ann. N.Y. Acad. Sci.* **448**, 361-374 (1985).
- Willeits, J., Urban, L., Murase, K. & Randic, M. *Ann. N.Y. Acad. Sci.* **448**, 385-401 (1985).
- Kikkawa, U., Takai, Y., Minakuchi, R., Inohara, S. & Nishizuka, Y. *J. biol. Chem.* **257**, 13341-13348 (1982).

Single stretch-activated ion channels in vascular endothelial cells as mechanotransducers?

Jeffrey B. Lansman*‡, Trevor J. Hallam† & Timothy J. Rink†

* Physiological Laboratory, Downing Street, University of Cambridge, Cambridge CB2 3EG, UK
† Smith Kline and French Research Limited, Welwyn, Hertfordshire AL6 9AR, UK

Endothelial cells line the inner surface of blood vessels and act as the main barrier to the passage of cells and large molecules from the blood stream to the tissues. Recent interest in the part played by the endothelium in regulating vascular tone has focused on the synthesis and secretion of prostacyclin^{1,2} and an endothelium-derived relaxing factor^{3,4}. Endothelial cells respond to blood-borne agonists⁵, but how the endothelium senses and responds to mechanical forces generated by the flow of blood under pressure is not known⁶⁻¹². Here we report patch-clamp recordings of ion channel activity from cell-attached membrane patches on aortic endothelial cells. In most of the patches examined, we observed unitary inward currents associated with the opening of a cation-selective channel (~40 pS in standard saline). The channel is permeable to Ca^{2+} and its opening frequency increases when the membrane is stretched by applying suction through the patch electrode. The presence of mechanotransducing ion channels¹³⁻¹⁵ in endothelial cells may help explain how the endothelium mediates vascular responses to haemodynamic stresses.

Figure 1 shows the pattern of single-channel activity recorded from a cell-attached patch, with physiological saline in both the patch electrode and in the bath. The records in Fig. 1a were obtained with the patch potential displaced -19 mV from the resting potential. Channel openings appeared as square pulses of inward current of about 1.4 pA. The relationship between the

holding potential and the amplitude of the unitary current is shown in Fig. 1b. The *I-V* relation was linear over the voltage range examined with a slope conductance of 39 ± 3.3 pS ($n = 7$).

An analysis of the kinetics of channel activity recorded at a constant voltage showed gating involved multiple open and closed states. At least two exponentials were needed to fit the open time histogram (Fig. 1c), corresponding to short and long openings in the single channel records (Fig. 1a). The distribution of closed times was not well resolved (Fig. 1d); there were at least two short-lived closed states corresponding to very brief gaps within long bursts and longer closings separating openings of both types. In addition, a much longer closed state having a time constant of seconds accounted for extended periods of silence between episodes of activity such as the one shown in Fig. 1a. There were too few of these very long closings to contribute a significant component to the histogram of closed times.

To show the changes in channel activity on a longer time scale, the number of openings in consecutive and non-overlapping time intervals was counted and plotted as a function of time as shown in Fig. 1e. Channel activity was high initially, but declined and remained low, except for the peak of activity in the interval marked with the bar (the records in Fig. 1a are from this period). The number of openings during this period was significantly greater than the expected mean calculated for the entire recording.

The rapid decline in channel activity after the electrode sealed to the membrane suggested channels might be responding to deformation of the membrane as it was sucked into the electrode as a seal was formed. Figure 2 shows the effect of suction applied by the electrode several minutes after seal formation. In this and subsequent experiments, the bathing solution contained isotonic potassium aspartate to set the cell resting potential to 0 mV (Fig. 2 legend). Three consecutive traces are shown with the patch potential held at -57 mV. When suction was applied through the electrode to the patch (indicated with arrows), openings of unitary amplitude occurred in rapid succession throughout the stimulus. We have not yet analysed the kinetics of the response to suction. Inspecting the current records suggests suction increased channel opening probability by increas-

‡ Present address: Department of Pharmacology, School of Medicine, University of California, San Francisco, California 94143, USA.

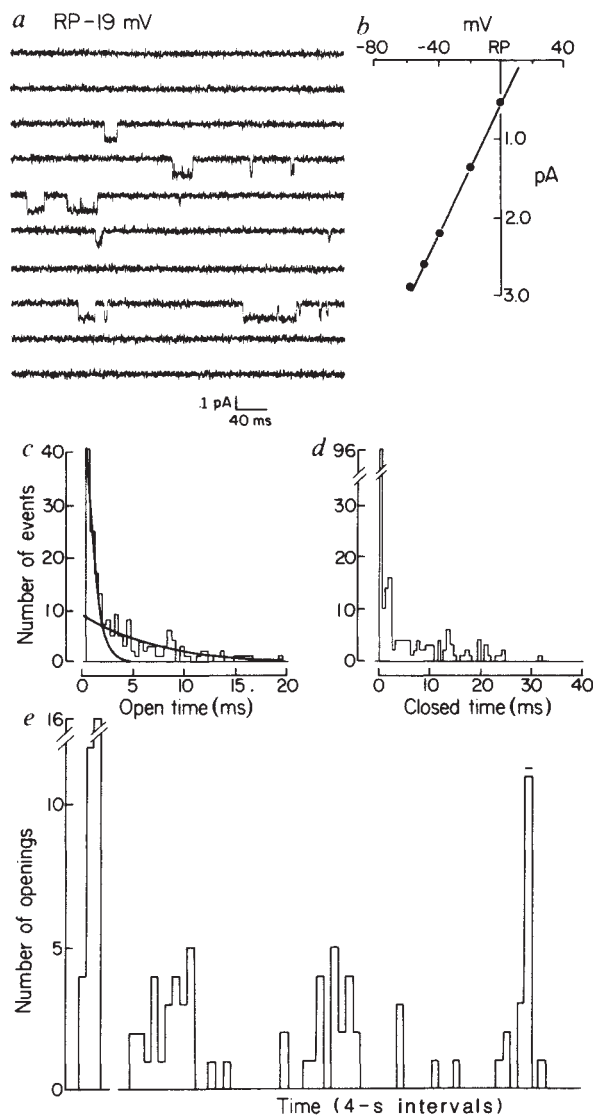


Fig. 1 Single-channel activity recorded from a cell-attached patch on an aortic endothelial cell. Both the bath and patch electrode contained standard saline (below). *a*, Spontaneous channel activity recorded from a cell-attached patch. Patch holding potential = resting potential -19 mV. *b*, Amplitude of the unitary event as a function of the applied potential. Although the resting membrane potential is unknown in this experiment, it was estimated to be about -30 mV by comparing the reversal with the cell was bathed in isotonic potassium aspartate and the membrane potential was known to be zero. *c*, *d*, Histograms of open and closed times, respectively. In the open time histogram in *c*, the fast exponential was drawn with a time constant of 0.7 ms, and the slower exponential with a time constant of 6.0 ms; both were fitted by eye. *e*, Channel activity during the first 5 min of the experiment. Zero time corresponds to the first records obtained after establishing a seal between the electrode and the membrane. The numbers of openings in consecutive and non-overlapping time intervals of 4.096 s were counted (discarding openings of two sample points or less) and plotted as a function of interval number. Records in *a* are from the interval marked with the horizontal bar. Cell 12-21-4. **Methods.** Endothelial cells were isolated from neonatal pig aorta and cultured as described²². Cells were passaged twice weekly. Only cells from the first 3–4 passages were used for experiments. Standard saline contained 150 mM NaCl, 5 mM KCl, 2.5 mM CaCl_2 , 1 mM MgCl_2 , 10 mM glucose, and 10 mM HEPES and was adjusted to pH 7.5 with NaOH. Current signals were recorded on magnetic tape. For analysis, currents were filtered with an eight-pole Bessel filter with a cut-off frequency (-3 dB) of 1 kHz and sampled at 5 kHz onto the hard disk of a PDP-11/23 laboratory computer. Experiments were done at room temperature ($\sim 21^\circ\text{C}$).

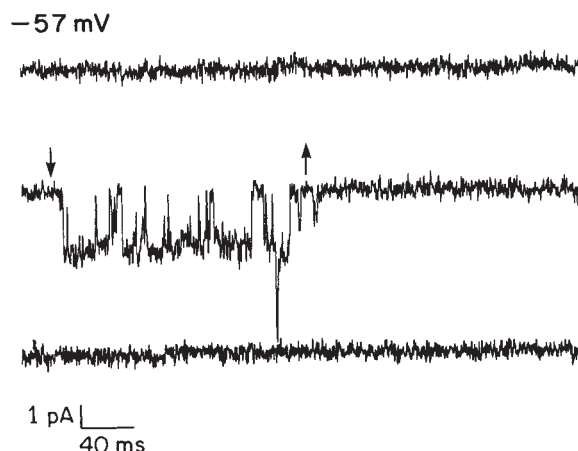


Fig. 2 Response of patch current to suction applied to the membrane through the electrode. In these experiments, suction pulses of 10 – 20 mm Hg (1.3 – 2.7 dyn cm^{-2}) were used to activate channel opening. Arrows, approximate onset and termination of the pulse of suction applied to the electrode through the suction line. The current records are from consecutive sweeps. The electrode contained standard saline. The bathing solution contained 150 mM KOH, 150 mM aspartic acid, 5 mM MgCl_2 , 1 mM EGTA, 10 mM glucose, and 10 mM HEPES adjusted to pH 7.5 with KOH. In the isotonic K^+ bathing solution the resting membrane potential is close to 0 mV, because excising the patch from the cell did not measurably alter the amplitude of the unitary current²³. Changing the bathing solution from physiological saline to isotonic potassium aspartate had no effect on channel activity or cell viability. Note the opening of a second channel of unitary amplitude. Holding potential was -57 mV. Cell 12-20-2.

ing the burst frequency, perhaps by shortening the duration of the longer closed periods between openings¹⁴.

Applying suction was the only stimulus we found that opened channels. Changing the membrane potential over a wide range failed to open channels in quiescent cells. Excising the patch into a bathing solution containing either 2 mM Ca^{2+} or 5 mM EGTA (no added Ca^{2+}) or including 2 mM Ca^{2+} in the electrode solution did not alter channel activity. On the other hand, suction had no effect on the activity of a 60 pS K^+ channel we have seen in other patches (data not shown).

We examined the permeability of the channel to various monovalent cations. Channels were identified by their response to suction, which was similar to that shown in Fig. 2. With isotonic KCl in the electrode, the channel conductance was 56.0 ± 11.0 pS and current reversed at -5.0 ± 3.5 mV ($n = 5$). With standard Na^+ -containing saline in the electrode, the reversal potential was $+8.6 \pm 16.0$ mV ($n = 5$). One experiment with isotonic CsCl in the electrode gave a conductance of 27 pS and a reversal potential of $+4.4$ mV. Replacing half the chloride with aspartate in the electrode solution or excising the patch so that its cytoplasmic surface was exposed to the isotonic potassium aspartate bath solution did not measurably alter either the conductance or the reversal potential, suggesting the channel is not permeable to chloride.

Figure 3 shows an experiment in which the patch electrode contained 110 mM CaCl_2 . Channel activity in response to suction consisted of a rapid succession of brief openings and closings (arrows); here the inward current is presumably carried by Ca^{2+} . The relationship between the patch potential and amplitude of the unitary current is shown in Fig. 3b. The slope conductance was 19.1 ± 3.0 pS and the reversal potential was 17 ± 13.0 mV ($n = 5$). Calculation of the relative permeability to Ca^{2+} and Na^+ from the reversal potential measurements with the constant field equation^{15–18} indicated the channel is somewhat more permeable to Ca^{2+} than to Na^+ ($P_{\text{Ca}^{2+}}/P_{\text{Na}^+} = 6.0$, range 1.2 – 8.4).

This paper reports the first measurements of ion-channel

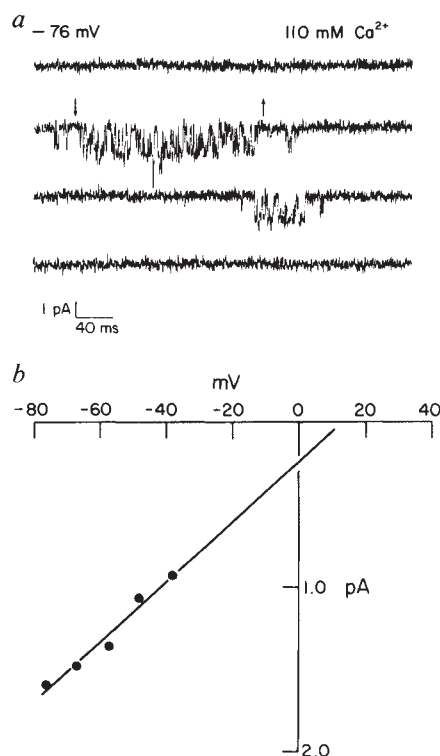


Fig. 3. Currents carried by Ca^{2+} through single stretch-activated channels. Records in *a* are from consecutive sweeps; arrows, approximate duration of the suction pulse. *b*, Amplitude of the single channel current plotted as a function of the applied voltage. Pipette solution contained 110 mM CaCl_2 , 10 mM glucose, and 10 mM HEPES adjusted to pH 7.5 with tetraethylammonium hydroxide. Cell resting potential zeroed with isotonic potassium aspartate. Holding potential was -76 mV. Cell 12-20-7.

activity in endothelial cells. The stretch-activated channel described here is similar to the channel in skeletal muscle described by Brehm *et al.*¹³ and Guharay and Sachs^{14,15}. Our conductance measurements agree with those reported by Guharay and Sachs, who also found the conductance to be larger in external K^+ than in Na^+ solutions. We found that the reversal potential is somewhat less sensitive to the species of monovalent cation. Our results also differ from those of Guharay and Sachs in that we find the histogram of open times has an additional slow component.

The results show that Ca^{2+} fluxes through stretch-activated channels in isotonic CaCl_2 are large enough for channel openings to be detected as unitary events. Although Na^+ would carry significant current under physiological conditions, sufficient Ca^{2+} may enter the cell to serve a second messenger function. Ca^{2+} mediates the synthesis and release of prostacyclin and endothelium-derived relaxing factor stimulated by various agonists^{3,4,19}. The fact that an intact endothelium is required for arterial contraction in response to stretching^{20,21} and for the vasodilation which occurs in response to shear stresses produced by the flow of blood over the endothelial surface^{8,12} makes it tempting to speculate that the effects of blood pressure and flow are mediated by mechanotransducers of the type described here.

We thank the Royal Society for an equipment grant, Dr Jeremy Pearson for help preparing cells, and Professor Denis Haydon for providing laboratory facilities. J.B.L. was supported by a NATO Fellowship.

- Peach, M. J., Loeb, A. L., Singer, H. A. & Saye, J. *Hypertension* **7**, 194-1100 (1985).
- Shepro, D. & D'Amore, P. A. in *Handbook of Physiology: The Cardiovascular System, Part 1* (eds Bohr, D. F., Somlyo, A. P. & Sparks, H. V.) 103-164 (American Physiological Society, Bethesda, 1984).
- DeForrest, J. M. & Hollis, T. M. *Am. J. Physiol.* **234**, H701-705 (1978).
- Franke, R.-P. *et al. Nature* **307**, 648-649 (1984).
- Holtz, J., Forstermann, U., Pohl, U., Giesler, M. & Bassenge, E. *J. Cardiovasc. Pharmacol.* **6**, 1161-1169 (1984).
- Frangos, J. A., Eskin, S. G., McIntire, L. V. & Ives, C. L. *Science* **227**, 1477-1479 (1985).
- Langille, B. L. & O'Donnell, F. *Science* **231**, 405-407 (1986).
- Davies, P. F., Remuzzi, A., Gordon, E. J., Dewey, C. F. & Gimbrone, M. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2114-2117 (1986).
- Pohl, U., Holtz, J., Busse, R. & Bassenge, E. *Hypertension* **8**, 37-44 (1986).
- Brehm, P., Kullberg, R. & Moody-Corbett, F. *J. Physiol., Lond.* **350**, 631-648 (1984).
- Guharay, F. & Sachs, F. *J. Physiol. Lond.* **352**, 685-701 (1984).
- Guharay, F. & Sachs, F. *J. Physiol. Lond.* **363**, 119-134 (1985).
- Jan, L. Y. & Jan, Y. N. *J. Physiol. Lond.* **262**, 215-236 (1976).
- Lewis, C. A. *J. Physiol., Lond.* **286**, 417-445 (1979).
- Edwards, C., Ottonson, D., Rydquist, B. & Swerup, C. *Neuroscience* **6**, 1455-1460 (1981).
- Weksler, B. B., Ley, C. W. & Jaffe, E. A. *J. clin. Invest.* **62**, 923-930 (1978).
- Katusic, Z. S., Shepherd, J. T., Vanhoutte, P. M. *Fedn. Proc.* **45**, 289 (1986).
- Hickey, K. A., Rubanyi, G., Paul, R. J. & Highsmith, R. F. *Am. J. Physiol.* **248**, C550-C556 (1985).
- Pearson, J. D., Carleton, J. S., Hutchings, A. & Gordon, J. L. *Biochem. J.* **170**, 265-271 (1978).
- Hess, P., Lansman, J. B. & Tsien, R. W. *J. gen. Physiol.* **88**, 293-319 (1986).

Relaxin affects the release of oxytocin and vasopressin from the neurohypophysis

Govindan Dayanithi*‡, Monique Cazalis* & Jean J. Nordmann†

*INSERM U.176, 33077 Bordeaux Cedex, France

†Centre de Neurochimie, 5, rue Blaise Pascal, 67084 Strasbourg Cedex, France

The hormone relaxin (for review see refs 1 and 2) has recently been shown to inhibit not only uterine muscle contraction³, but also the release of oxytocin into the plasma. Intravenous injection of porcine relaxin in anaesthetized lactating rats inhibits milk ejection and injection of relaxin into the cerebral ventricles disturbs the pattern of the milk ejection reflex⁴. Recent experiments performed *in vivo* indicate that relaxin might act not only in the uterus, but also in the hypothalamus and possibly in the neurohypophysis⁵. We tested this hypothesis *in vitro* by studying the effect of relaxin on hormone release from isolated neural lobes of the pituitary and isolated neurosecretory nerve endings of the neurohypophysis from the rat. We report here that relaxin has a dual effect on neurohypophysial hormone secretion. Under basal conditions, vasopressin and oxytocin release was inhibited by relaxin but, when the nerve endings were depolarized, vasopressin and oxytocin secretion was potentiated. We also found that relaxin acts at a stage before the increase in cytoplasmic free Ca^{2+} that is necessary for inducing hormone release, possibly by gating the calcium channel.

Figure 1 illustrates the time course of vasopressin (AVP) and oxytocin (OT) release from isolated nerve terminals in the presence or absence of 10^{-7} M relaxin. An increase in hormone secretion was induced by depolarizing the isolated nerve terminals (neurosecretosomes) with 100 mM K^+ for 10 min. Relaxin decreased the basal AVP and OT release by 53% and 43% respectively (Table 1). On the other hand, relaxin greatly potentiated the K^+ -evoked hormone release. The AVP release increased by 98% and that of OT by 59% (Table 1).

Figure 2 shows the dose-response curves of K^+ -evoked (Fig. 2a) and basal (Fig. 2b) AVP and OT release obtained from experiments similar to those described in Fig. 1. For the evoked hormone release, half-maximal activation of AVP and OT release occurred at concentrations of 13 nM and 25 nM relaxin respectively. For release under resting conditions, the K_D was 1 nM for AVP and 2.5 nM for OT. These values might be overestimated, because relaxin binds to plastic material such as

Received 28 July; accepted 18 December 1986.

- Weksler, B. B., Marcus, A. J. & Jaffe, E. A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3922-3926 (1977).
- Marcus, A. J., Weksler, B. B. & Jaffe, E. A. *J. biol. Chem.* **253**, 7138-7141 (1978).
- Furchgott, R. F. *Circulation Res.* **53**, 557-573 (1983).

‡ Present address: Centre de Neurochimie, 67084 Strasbourg Cedex, France.

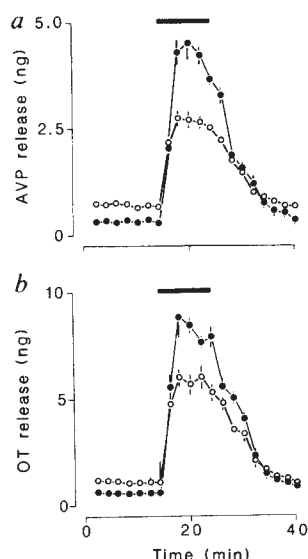


Fig. 1 Effects of relaxin on oxytocin and vasopressin release from isolated neurosecretory nerve terminals stimulated with elevated potassium concentration. The nerve endings were perfused with saline in the presence of 10^{-7} M relaxin. Stimulation of release was induced, during the time indicated by the heavy bar, with 100 mM potassium, choline being reduced accordingly. The amounts of vasopressin (a) and oxytocin (b) released in the perfusate were measured by radioimmunoassay. ●, Amounts released in the presence of relaxin; ○, controls. Note that under basal conditions relaxin decreases the amount of hormone released. The results are the mean $\pm$ s.e.m. of 4 experiments. Standard errors larger than the symbols used are shown as bars.

Methods. The isolated nerve endings were prepared by homogenizing two neural lobes isolated from male Wistar rats (body weight 250–300 g) in 1 ml of 0.27 M sucrose containing 10 μ M EGTA and buffered with 10 mM HEPES at pH 7.2 (ref. 9). The homogenate was centrifuged at 600g for 4 min and the resulting supernatant was centrifuged at 3,400g for 15 min. The pellet contained mostly isolated nerve endings and swellings as checked by electron microscopy. The pellet was resuspended in normal saline containing (mM): NaCl, 140; KCl, 5; NaHCO₃, 5; CaCl₂, 2.2; MgCl₂, 1; glucose, 10; HEPES, 10; pH 7.2 and to which BSA, at a final concentration of 0.01%, was added. The nerve endings were loaded onto a 0.22 μ M Millipore filter and perfused with saline for 35 min at a flow rate of 50 μ l min⁻¹. The neurosecretosomes were then perfused with Na-free saline containing an equimolar quantity of choline and the flow rate was increased to 100 μ l min⁻¹. Collection of the perfusate started 45 min after loading.

that used in the perfusion system (G. Bryant-Greenwood and C. Bagnell, personal communication). Thus, *in vivo*, relaxin might modulate the release of neurohypophysial hormones at even lower concentrations.

In another series of experiments, isolated neural lobes, separated from the pars intermedia, were electrically stimulated with a pattern of firing mimicking the electrical activity of the mag-

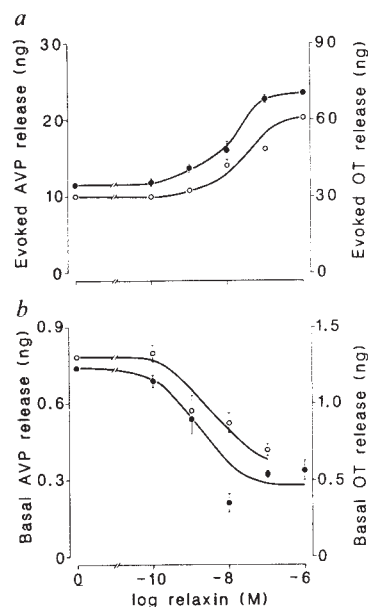


Fig. 2 The effects of increasing concentrations of relaxin on evoked and basal oxytocin and vasopressin release from isolated nerve terminals. The neurosecretosomes were perfused with Na-free saline and depolarized with 100 mM K⁺ as in Fig. 1. The evoked release (a) of (●) AVP and (○) oxytocin was calculated by subtracting the mean basal release (b), determined in the fractions preceding the onset of the stimulus, from the amount of hormone found in each fraction. Relaxin, at the different concentrations indicated on the abscissa, was present before and during the K⁺ stimulation. Values are mean $\pm$ s.e.m. ($n=4$).

nocellular neurons of the hypothalamus⁶. Relaxin (10^{-7} M) potentiated both AVP and OT release when hormone secretion was triggered either with AVP-like bursts (Fig. 3; Table 1) or with electrical discharges mimicking the activity of oxytocin-containing cells during lactation (Table 1). In the presence of relaxin, the firing pattern of vasopressin-containing cells potentiated the release of AVP by 119% and that of OT by 79% (Table 1). Potentiation of release by relaxin also occurred following stimulation with OT-like discharges. At a relaxin concentration of 10^{-7} M, AVP and OT secretion was enhanced by 88% and 69% respectively (Table 1). Finally, we tested the effect of relaxin on hormone release from isolated neural lobes stimulated with a pattern mimicking the electrical activity of a cell with a slow firing rate. This type of firing can be observed in cells under resting conditions⁷. The mean firing rate was only 3.9 Hz, but the discharges were made up of small bursts containing four spikes separated by intervals of 15 ms. Thus, although the mean firing rate was 3.9 Hz the frequency of firing within the burst was much higher. The neural lobes were stimulated for 20 min and this gave rise to an increased hormone release which was, for AVP and OT, potentiated in the presence of relaxin (10^{-7} M) by 132% and 69% respectively (Table 1). Furthermore, as shown

Table 1 The effects of porcine relaxin on release of vasopressin and oxytocin from isolated nerve terminals and neural lobes

	Vasopressin		Oxytocin	
	Control	+Relaxin (10^{-7} M)	Control	+Relaxin (10^{-7} M)
Nerve endings:				
Basal	0.37 $\pm$ 0.02 (4)	0.17 $\pm$ 0.01 (4)	0.61 $\pm$ 0.07 (4)	0.35 $\pm$ 0.04 (4)
K ⁺ -induced secretion	11.50 $\pm$ 0.50 (4)	22.70 $\pm$ 0.60 (4)	29.90 $\pm$ 1.00 (4)	47.40 $\pm$ 1.30 (4)
Neural lobe:				
Basal	0.11 $\pm$ 0.01 (18)	0.07 $\pm$ 0.003 (15)	0.30 $\pm$ 0.03 (18)	0.08 $\pm$ 0.005 (15)
'AVP-like' bursts	2.75 $\pm$ 0.14 (5)	5.54 $\pm$ 0.14 (5)	5.00 $\pm$ 0.18 (5)	8.94 $\pm$ 0.18 (5)
'OT-like' discharges	0.56 $\pm$ 0.06 (5)	1.05 $\pm$ 0.07 (5)	2.47 $\pm$ 0.17 (5)	4.31 $\pm$ 0.08 (4)
'Slow-firing' cell	1.07 $\pm$ 0.06 (8)	2.48 $\pm$ 0.10 (8)	1.78 $\pm$ 0.09 (8)	3.00 $\pm$ 0.12 (5)

Basal secretion and evoked hormone release are expressed as ng min⁻¹ and ng respectively. The results are given as mean $\pm$ s.e.m. with the number of experiments indicated in parentheses.

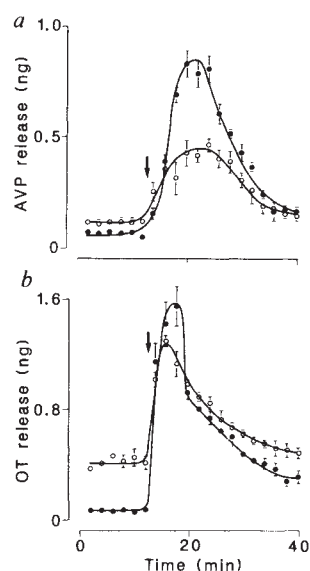


Fig. 3 Effect of relaxin on oxytocin and vasopressin release from electrically stimulated rat neural lobe. The neural lobes were stimulated with burst patterns mimicking the electrical activity of vasopressin-containing cells during haemorrhage. Arrows, onset of stimulation. AVP, (a); and OT, (b), in the perfusate were measured by radioimmunoassays. Values are the mean $\pm$ s.e.m. ($n=5$).

Methods. The neural lobes were obtained from male rats (body weight 250–300g) and the pars intermedia was carefully dissected out⁹. The neural lobe was impaled on one of a pair of electrodes which were inserted in a small perspex chamber which had an internal volume of about 50 μ l. The gland was continuously perfused with normal saline (see Fig. 1 legend) in the presence or absence of relaxin (10^{-7} M). Fractions of 2 min duration (100μ l min^{-1}) were collected 45 min after the onset of the perfusion. The neurohypophyses were stimulated electrically with biphasic electrical pulses which had a pattern of discharge similar to the bursting activity of AVP cells⁶. Briefly, the electrical activity of a vasopressin cell recorded *in vivo* was used as a command pulse to trigger the stimulator and thus generate a series of pulses (2 ms duration, 4 mA, biphasic) similar to the bursting pattern observed *in vivo*. The neural lobes were stimulated with four 'AVP-like' bursts separated by intervals of 21 s.

above on the isolated nerve endings, relaxin considerably decreased AVP and OT secretion from the neurohypophysis under resting conditions (Table 1).

Naloxone, the function of which in the neurohypophysial release mechanism is still very controversial^{8,9}, did not modify the potentiating effect of relaxin. In the presence of 10^{-7} M relaxin, the 100 mM K^+ -evoked hormone release from the neurosecretosomes was, in the presence and absence of 10^{-5} M naloxone, 24.3 ± 0.3 and 22.7 ± 0.6 ng for AVP, and 44.1 ± 2.1 and 47.4 ± 1.3 ng for OT. Furthermore, naloxone did not prevent the relaxin-induced inhibition of AVP and OT release (data not shown). This shows that relaxin does not interact, at least in the neurohypophysis, with a naloxone-sensitive opiate receptor.

To see if relaxin acts after the increase in ionized cytoplasmic concentration induced by the activation of voltage-dependent calcium channels, it was added before and during a period when internal calcium was artificially increased without depolarising the membrane. This involves making the neurosecretosomes permeable with digitonin in a medium containing micromolar concentrations of calcium¹⁰. Briefly, the nerve endings were perfused with normal saline, then with Ca-free medium of the following composition (mM): potassium glutamate, 130; MgCl_2 , 2; EGTA, 2; glucose, 10; PIPES, 20. The medium contained 0.01% bovine serum albumin (BSA) and its pH was adjusted to 6.8. The neurosecretosomes were permeabilized, for a period of 10 min, with $1 \mu\text{M}$ digitonin in the presence of $1.1 \mu\text{M}$ Ca and 1 mM ATP. The Ca-dependent hormone release was calcu-

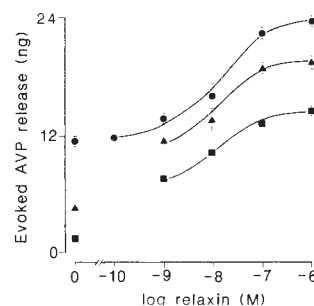


Fig. 4 Effects of relaxin on the depolarization-induced AVP release from nerve endings in the presence of nitrendipine. The neurosecretosomes were prepared as in Fig. 1 and perfused with different concentrations of relaxin. Hormone release was triggered with 100 mM K^+ . Relaxin (10^{-7} M) and nitrendipine (4×10^{-6} M, $\blacktriangle$ or 15×10^{-6} M, $\blacksquare$) were present throughout the experiments. $\bullet$, Evoked AVP release in the absence of nitrendipine. Values are mean $\pm$ s.e.m. ($n=4$).

lated by subtracting the amount released in the presence of calcium from that secreted after permeabilization in a Ca-free medium. The Ca-independent release corresponded to $<10\%$ of the total amount secreted, suggesting that, after permeabilization, only a small amount of hormone was released by a non-specific mechanism. Relaxin did not modify the total evoked release of AVP and OT. In controls, the AVP and OT levels were 3.20 ± 0.07 ng ($n=4$) and 6.20 ± 0.77 ng ($n=4$) respectively. In the presence of relaxin (10^{-7} M) these values were 3.06 ± 0.07 ng ($n=4$) and 5.84 ± 0.27 ng ($n=4$) respectively. Thus relaxin does not act at a stage which couples the increased ionized cytoplasmic calcium concentration to the mechanism of exocytosis *per se* but interferes with the Ca entry preceding the release of hormone. Figure 4 shows the effects of different concentrations of relaxin on AVP release from neurosecretosomes in the presence of the calcium channel blocking agent nitrendipine. This dihydropyridine molecule is a potent antagonist of the depolarization-induced hormone release from isolated neurohypophysial nerve endings¹¹. From its action on calcium current in other preparations^{12–14}, it can reasonably be assumed that it blocks voltage-gated calcium channels in the neurosecretory nerve endings. Relaxin considerably inhibits the effect of nitrendipine on AVP and OT release. Experiments performed with different relaxin and nitrendipine concentrations (Fig. 4) suggest a non-competitive inhibition by relaxin of the effect of nitrendipine on AVP release. Figure 4 also shows that nitrendipine did not affect the relaxin concentration necessary to induce half-maximal activation of hormone release but decreased, in a dose-dependent manner, the amount of AVP secreted after K^+ -induction. We do not yet know if relaxin acts directly on the calcium channel at a binding site distinct from that of nitrendipine, or at a site associated but distinct from the calcium channel. Patch-clamp studies on neurosecretosomes^{15,16} should help to answer this question soon. The dual effect of relaxin on basal and evoked hormone release makes it an attractive candidate for modulating the gating mechanism of the calcium channels in the plasma membrane of neurosecretory nerve terminals.

In conclusion, the present work shows that relaxin, a hormone produced *in vivo* by the ovary, can directly modulate the secretion of neuropeptides from the neural lobe of the pituitary. To our knowledge, this is the first direct evidence that a circulating hormone controls neuropeptide secretion from neurohypophysial nerve terminals. It is worth noting that relaxin decreases the milk ejection reflex when injected into anaesthetized lactating rats, presumably by acting within the central nervous system⁴. Further electrophysiological studies are needed to determine the effect of relaxin on the electrical activity of the magnocellular neurons. We do not yet know if the relaxin implicated in the release of neurohypophysial hormones is synthesized at the periphery or in the hypothalamus. The results

demonstrate that relaxin acts at a step in the cascade leading to exocytosis which precedes the increase in free cytoplasmic Ca^{2+} concentration induced by depolarization of the plasma membrane. We suggest that in neurohypophysial nerve endings relaxin directly modulates the entry of calcium through Ca channels.

We thank G. Bryant-Greenwood and C. Bagnell for their gift of relaxin and for helpful discussion. We acknowledge the participation of M. Y. Flute and B. A. S. Kinrobbins in preliminary stages of this work. We also thank D. Poulain for recording the magnocellular neurons and providing us with his computer program, and J. P. Lemos for help with the manuscript. The antisera for AVP and OT were gifts from Drs P. Czernikow and A. P. F. Flint respectively. G.D. is a recipient of a French Government Fellowship. Supported by grants from CNRS and INSERM.

Received 16 October; accepted 30 December 1986.

- Porter, D. G. in *Oxford Review of Reproduction and Endocrinology*, Vol. 1 (ed. Finn, C. A.) 1-57 (Clarendon Press, Oxford, 1979).
- Bryant-Greenwood, G. D. *Endocr. Rev.* 3, 62-90 (1982).
- Bradshaw, J. M. C., Downing, S. J., Moffatt, A., Hinton, J. C. & Porter, D. G. *J. Reprod. Fert.* 63, 142-153 (1981).
- Summerlee, A. J. S., O'Byrne, K. T., Paisley, A. C., Breeze, M. F. & Porter, D. G. *Nature* 309, 372-374 (1984).
- O'Byrne, K. T. & Summerlee, A. J. S. *J. Physiol., Lond.* 365, 49P (1985).
- Cazalis, M., Dayanithi, G. & Nordmann, J. J. *J. Physiol., Lond.* 369, 45-60 (1985).
- Poulain, D. & Wakerley, J. B., *Neuroscience* 7, 773-808 (1982).
- Bicknell, R. J. *J. Endocr.* 107, 437-446 (1985).
- Nordmann, J. J., Dayanithi, G. & Cazalis, M. *Expl Brain Res.* 61, 560-566 (1986).
- Cazalis, M., Dayanithi, G. & Nordmann, J. J. *J. Physiol., Lond.* (in the press).
- Cazalis, M., Dayanithi, G. & Nordmann, J. J. *J. Physiol., Lond.* (in the press).
- Freedman, S. B., Dawson, G., Villeral, M. L. & Miller, R. J. *J. Neurosci.* 4, 1453-1467 (1984).
- Hess, P., Lansmann, J. B. & Tsien, R. W. *Nature* 311, 538-544 (1985).
- Reuter, H., Porzig, H., Kokubun, S. & Prod'homme, B. *Trends Neurosci.* 8, 396-400 (1985).
- Lemos, J. R., Nordmann, J. J., Cooke, I. M. & Stuenkel, E. L. *Nature* 319, 410-412 (1986).
- Lemos, J. R. & Nordmann, J. J. *J. exp. Biol.* 124, 53-72 (1986).

Redesigning the body plan of *Drosophila* by ectopic expression of the homoeotic gene *Antennapedia*

Stephan Schneuwly, Roman Klemenz
& Walter J. Gehring

Department of Cell Biology, Biozentrum, University of Basel,
Klingelbergstr. 70, CH-4056 Basel, Switzerland

Genetic and molecular studies on the expression of *Antennapedia* (*Antp*) have suggested that this gene specifies mainly the second thoracic segment. On the basis of our molecular analysis of dominant gain-of-function mutants we have postulated that the transformation of antennae into second legs is due to the ectopic overexpression of the *Antp*⁺ protein. This hypothesis was tested by inserting the complementary DNA encoding the normal *Antp* protein into a heat-shock expression vector and subsequent germline transformation. As predicted, heat induction at defined larval stages leads to the transformation of antennae into second legs. The dorsal part of the head can also be transformed into second thoracic structures (scutum) indicating that *Antp* indeed specifies the second thoracic segment. By ectopic overexpression of the *Antp* protein the body plan of the fruit fly can be altered in a predictable way.

Mutations at the homoeotic *Antennapedia* (*Antp*) locus affect the development of all three thoracic segments of *Drosophila*¹⁻⁴. Recessive loss-of-function mutants, which correspond to the deletion of the locus, are late embryonic or larval lethals in which mainly the mesothoracic segment (T2) is transformed towards a prothoracic segment (T1) and head segments. The metathoracic segment (T3) is also partially transformed towards the meso- or prothoracic segment, that is in an anterior direction. Localization of *Antp*⁺ transcripts by *in situ* hybridization^{5,6} and

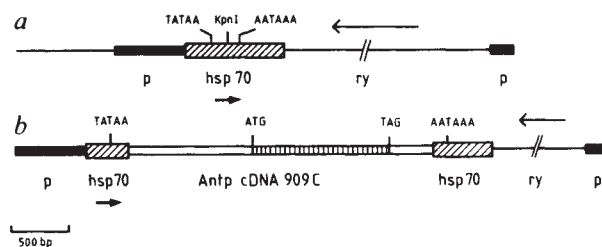


Fig. 1 Structure of the heat shock *Antp* fusion plasmids. **a**, Vector plasmid pHT4. This vector was constructed by inserting the *hsp70* promoter and the termination-processing sequences into Carnegie 20²⁴, a vector which contains the *rosy* (*ry*⁺) gene flanked by the ends of the P-transposon²⁵. The unique *KpnI* site separates the *hsp70* leader from the *hsp70* trailer sequences. This site is suitable for cloning of passenger sequences, resulting in a fusion to the *hsp70* leader and expression under the heat-shock promoter. The leader fragment contains the first 95 base pairs (bp) of the *hsp70* leader and 250 bp upstream sequences. The trailer fragment contains the whole trailer sequence of *hsp70* and 350 bp of non transcribed 3' flanking sequences. **b**, Heat shock *Antp* fusion plasmid pHTA. The 2.8-kb *EcoRI* fragment from *Antp* cDNA 909C^{12,13} containing the whole open reading frame was made blunt ended with the Klenow fragment of *E. coli* DNA polymerase I and cloned into vector pHT4, which was linearized with *KpnI* and made blunt ended by a T4 polymerase reaction.

of the *Antp*⁺ protein by immunofluorescence⁷ is consistent with the idea that *Antp*⁺ mainly specifies the mesothoracic segment (T2), and also parts of T1 and T3, as both the RNA and the protein accumulate in these segments of the wild-type embryo. There is an even more striking class of mutants which are dominant and show a transformation of the adult antennae into second legs, towards the mesothorax, that is in the opposite direction to recessive loss-of-function mutants. These mutations are thought to represent a dominant gain-of-function, because they are dominant in triploids and their dominant phenotype can be 'reverted' by deleting the *Antp* gene^{8,9}. An exceptional dominant gain-of-function mutant *Cephalothorax* (*Ctx*)¹⁰ transforms the posterior dorsal part of the head into dorsal mesothoracic structures. However, this mutation results from a very complex chromosomal rearrangement with one breakpoint in the *Antp* locus¹¹ and only a single allele is known. Most of the mutations that lead to a transformation of antennae into second legs are chromosomal inversions^{11,12}. The structural analysis of the *Antp* gene and the determination of the nucleotide sequence of the relevant parts¹³ show that in all of the dominant gain-of-function mutants which have been analysed, the protein-coding region of *Antp* is not altered, but separated from its control region. Therefore, we have proposed that the dominant phenotype is not due to an altered gene product but is due rather to altered expression of the normal *Antp* protein. At least for the inversion *In(3R)Antp*^{73b}, we have been able to show that the chromosomal inversion results in the fusion of a 5' control region and RNA coding sequences of a foreign gene to the *Antp*⁺ protein-coding region giving rise to fusion transcripts¹⁴. As at least eight independent inversions are known¹¹⁻¹³ to produce antenna-to-leg transformations it seemed *a priori* unlikely that all these inversion events are due to fusion of the *Antp* protein-coding region to promoters which are expressed specifically in the antennal imaginal disk. This led to the speculation that the dominant gain-of-function phenotype may be caused by the ectopic overexpression of the *Antp* protein under the control of any promoter that would be expressed in the antennal imaginal disk at the appropriate time in development. Such a promoter might also be active in other tissues, but the effect would be limited to the antennae if the protein were expressed at the critical time when antennae are determined.

To test this hypothesis of ectopic overexpression we constructed an artificial gene fusion in which the *Antp* protein-coding region is fused to the heat-inducible promoter of the heat-shock

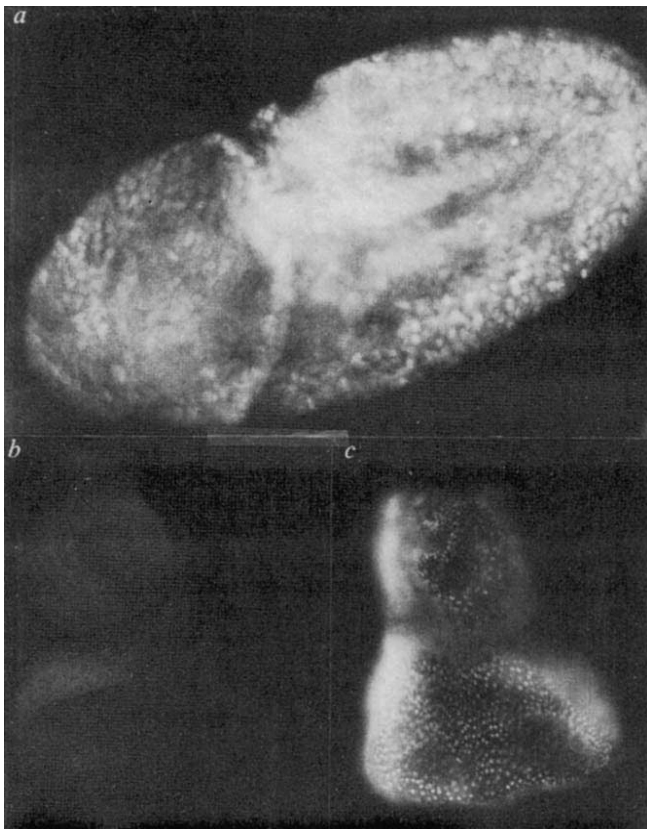


Fig. 2 *Antp* protein expression after heat induction. *a*, Embryos of transformant H4 were incubated for 7 h at 25 °C, heat shocked for 4 h at 37 °C, fixed, stained with anti-*Antp* antiserum⁷ and mounted in toto. Note that the entire embryo is labelled and that the antigen is located in the nuclei. *b, c*, Eye-antennal disks of late third-instar larvae from the transformant H4 were fixed and stained as above, with and without heat induction. In the non-heat-shock control disk (*b*) only cytoplasmic background staining is detected, whereas all nuclei are labelled after a 2 h heat shock at 37 °C followed by a 2 h recovery period at 25 °C.

gene *hsp70*¹⁵⁻¹⁷. This fusion gene has been introduced into the germline of wild-type recipient flies by P-element mediated germline transformation. As the *Antp* gene is very large and comprises more than 100 kilobases (kb), an *Antp* cDNA was inserted between the heat-shock promoter *hsp70* and its termination-processing sequences. For this purpose the *Antp* cDNA (909c) which contains the entire large open reading frame¹³ was inserted into the *Kpn*I site of the newly constructed pHT4 vector (Fig. 1*a*). The final construct, plasmid pHTA, used for germline transformation is shown in Fig. 1*b*. DNA of pHTA was coinjected with p π 25.7 w.c. helper plasmid into *ry*⁵⁰⁶ (ORM) recipient eggs¹⁸. Several independent transformed *ry*⁺ fly lines were established. They have insertions on different chromosomes (lines H4, H26, H49 on the 3rd, H22, H36 on the 2nd and H21, H45 on the X-chromosome) and all of them are homozygous viable except for line H21, in which the insertion seems to have induced a lethal mutation. Line H4 was used for most of the experiments described below. Other lines showed similar results both in terms of protein accumulation and developmental response.

To determine whether the integrated fusion gene produced *Antp* protein under the control of the heat-shock promoter, the localization of the protein was monitored by immunofluorescence labelling with a polyclonal mouse anti-*Antp* serum. This antiserum, which was obtained by immunizing mice with a protein synthesized in *Escherichia coli* that contains parts of the *Antp* open reading frame, recognizes a nuclear antigen in late

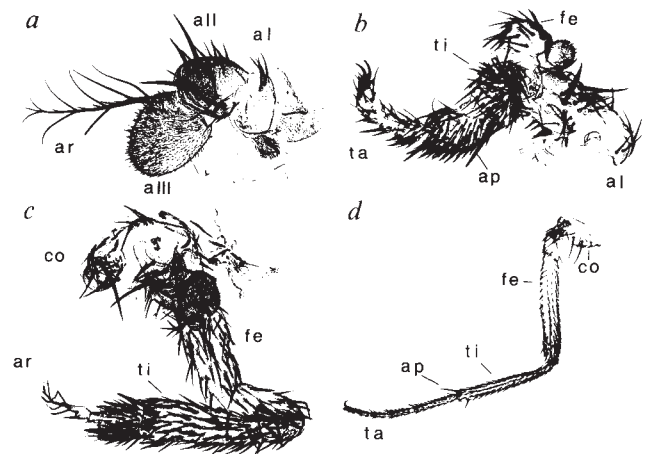


Fig. 3 Induction of *Antp* phenotype by heat shock. *a*, Normal antenna ($\times 68$). *b, c*, Two examples of antennae from heat-shocked flies. Heat shocks were for 2 h at 37 °C on the 4th and 5th days of development (3rd-instar larvae). The main parts of the antennae are transformed into leg structures. In *b*, the arista is completely transformed into tarsal structures including claws, whereas in *c* more proximal structures are transformed. The presence of the apical bristle indicates the mesothoracic identity of the leg ($\times 22$) ($\times 68$). *d*, Normal mesothoracic leg ($\times 22$). Abbreviations: aI, antennal segment I; aII, antennal segment II; aIII, antennal segment III; ar, arista; co, coxa; fe, femur; ti, tibia; ap, apical bristle; ta, tarsus.

embryos which is mainly expressed in the three thoracic segments of the epidermis and the nervous system (T2, parts of T1 and T3) and corresponds to the *Antp*⁺ protein⁷. Following exposure of transformant embryos to heat shock, *Antp*⁺ protein can be detected in all nuclei of the embryo (Fig. 2). The nuclear localization was confirmed by double labelling with a nuclear dye (Hoechst 33258). These results demonstrate that these transformants produce an *Antp* protein which, after heat shock, is expressed throughout the embryo. The same result was obtained with larvae. Closer examination of the eye-antennal disks of transformed larvae (Fig. 2*b* and *c*) indicates that in the heat-shocked disks all nuclei are strongly labelled by the antiserum, whereas in the non-heat-shocked controls only weak non-specific staining of the cytoplasm can be detected.

The developmental effects of heat induction were tested by temperature shift experiments. At the normal temperature (25 °C) the transformed flies showed no obvious phenotypic alterations although the *hsp70* promoter is known to be weakly active even at this temperature¹⁹. The effect of heat induction of *Antp* was examined by applying a heat shock at defined developmental stages. Eggs were collected in small tubes at 25 °C and allowed to develop at this temperature. A 1–2 h heat shock at 37 °C was applied at various stages and the tubes were returned to 25 °C to allow continued development. Figure 3 shows the results of experiments in which the transformed flies were exposed to heat shock during the third larval stage as described in the legend. In more than 50% of the heat-shocked flies, the antennae were clearly transformed into leg structures. On the basis of the bristle pattern, these leg structures correspond to those of the mesothoracic leg as indicated for example by the characteristic apical bristle on the tibia. In most cases the transformation is incomplete, but it corresponds exactly to those found in strong *Antp* mutant alleles. Therefore, heat-induced ectopic expression of the *Antp*⁺ protein results in the formation of mesothoracic legs in place of antennae, that is in an alteration of the body plan as predicted by our hypothesis. No such transformations have been observed in control flies which were transformed with the vector pHT 4 alone without the *Antp* insert and heat shocked at the same stages of development.

Closer examination of the adult phenotype following heat shock revealed additional transformations. As shown in Fig. 4*a*

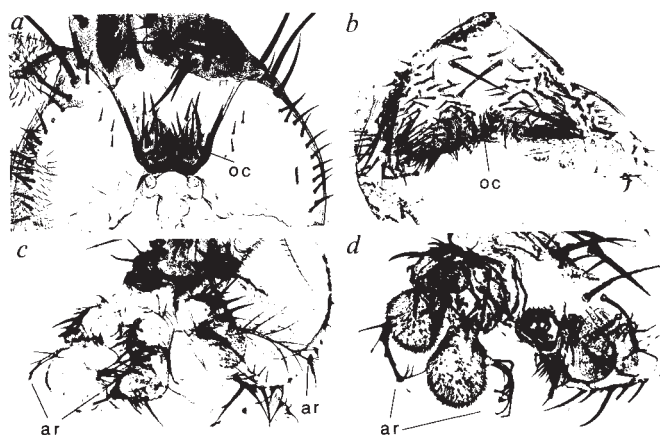


Fig. 4 Induction of *Cephalothorax* and *Dfd*-like phenotypes. *a*, Normal hind head with occiput ($\times 52$). *b*, Hind head from a fly, which was treated in the same way as in Fig. 3. Note the large symmetrical outgrowth of thoracic structures which resembles the *Cephalothorax* phenotype ($\times 52$). *c*, *d*, *Dfd*-like phenotype after heat shock during 1st and 2nd instar larval stage. In *c*, a complete additional antenna replaces one eye whereas in *d* parts of an antenna are duplicated as indicated by the presence of two aristae ($\times 45$). Abbreviations: ar, arista; oc, occiput.

and *b*, a region of the hindhead of heat-shocked flies is transformed into dorsal mesothoracic structures of the scutum. This phenotype corresponds to a strong *Ctx* phenotype and presumably represents another gain-of-function phenotype. It occurs at about the same frequency as the antenna-to-leg transformations. Other head defects and exceptional cases of antennal duplication have also been noted (Fig. 4c and d). These phenotypic effects are observed rarely. They resemble flies homozygous for the recessive *Deformed* allele (*Dfd*^{r-L}).²⁰

The precise developmental time during which the antenna-to-leg transformation can be induced is critical. Heat shock during the first and second larval instar produces no apparent effects and if the heat shock is applied late in development, just before pupation, only a few leg bristles with their characteristic bract are induced. A heat shock during precisely defined embryonic stages results in complete lethality of the transformants. Such embryos fail to involute the head segments properly and also exhibit segmental transformations towards T2 that will be described elsewhere (G. Gibson, S.S. & W.J.G., in preparation).

These transformation experiments demonstrate that the earlier gene assignment based on various *Antp* mutants^{11,12} was correct and that the 909c open reading frame encodes a functional *Antp* protein. They also show conclusively that *Antp* controls the developmental pathway leading to the second thoracic segment and specifies not only the ventral parts (the mesothoracic legs) but also dorsal parts (the scutum). We interpret the observation that ectopic expression of *Antp* protein can lead to a homeotic segmental transformation in terms of a model which assumes the competitive regulatory interaction among the various homeotic genes²¹. As shown by earlier experiments on trans-determination^{22,23}, the antennal disk seems to be a relatively weak spot in the genetic circuitry as a homeotic transformation from antennae to leg can occur spontaneously during *in vivo* culture. After deciphering the genetic circuits that control development further, it should be possible to understand the principles underlying the specification of the body plan both in molecular and evolutionary terms.

We thank G. Rubin and A. Spradling for plasmids p π 25.7w.c. and car20, Y. Hiromi and L. Pick for critical comments on the manuscript, D. Wey and U. Kloter for their help in preparing photographs and E. Wenger-Marquardt for efficient typing of the manuscript. This work was supported by grants from the Swiss National Science Foundation and the Cantons of Basel.

Received 28 July 1986; accepted 7 January 1987.

1. Wakimoto, B. T. & Kaufman, T. C. *Dev Biol.* **81**, 51–64 (1981).
2. Struhl, G. *Nature* **292**, 635–638 (1981).
3. Schneuwly, S. & Gehring, W. J. *Dev Biol.* **108**, 377–386 (1985).
4. Sato, T., Hayes, P. H. & Denell, R. E. *Dev Biol.* **111**, 171–182 (1985).
5. Levine, M., Hafen, E., Garber, R. L. & Gehring, W. J. *EMBO J.* **2**, 2037–2046 (1983).
6. Martinez-Arias, A. *EMBO J.* **5**, 135–141 (1986).
7. Wirz, J., Fessler, L. I. & Gehring, W. J. *EMBO J.* **5**, 3327–3334 (1986).
8. Denell, R. E. *Mutat Res.* **15**, 221–223 (1972).
9. Duncan, I. W. & Kaufman, T. C. *Genetics* **80**, 733–752 (1975).
10. Duncan, I. W. & Lewis, E. B. in *Developmental Order: Its Origin and Regulation* (eds Subtelny, S. & Green, P. B.) 533–554 (Liss, New York, 1982).
11. Scott, M. P. *et al. Cell* **35**, 763–776 (1983).
12. Garber, R. L., Kuroiwa, A. & Gehring, W. J. *EMBO J.* **2**, 2027–2036 (1983).
13. Schneuwly, S., Kuroiwa, A., Baumgartner, P. & Gehring, W. J. *EMBO J.* **5**, 733–739 (1986).
14. Schneuwly, S., Kuroiwa, A. & Gehring, W. J. *EMBO J.* (in the press).
15. Pelham, H. R. B. *Cell* **30**, 517–528 (1982).
16. Lis, J. T., Simon, J. A. & Sutton, C. A. *Cell* **35**, 403–410 (1983).
17. Craig, E. A. *CRC Crit. Rev. Biochem.* **18**, 239–290 (1985).
18. Hiromi, Y., Kuroiwa, A. & Gehring, W. J. *Cell* **43**, 607–617 (1985).
19. Steller, H. & Pirota, V. *EMBO J.* **4**, 3765–3772 (1985).
20. Vogt, M. *Biol. Zbl.* **66**, 81–105 (1947).
21. Gehring, W. J. *Science* (in the press).
22. Gehring, W. J. *J. Embryol. exp. Morph.* **15**, 77–111 (1966).
23. Gehring, W. J. *Dev Biol.* **16**, 438–456 (1967).
24. Rubin, G. M. & Spradling, A. C. *Nucleic Acids Res.* **11**, 6341–6351 (1983).
25. Rubin, G. M. & Spradling, A. C. *Science* **218**, 348–353 (1982).

Homology between *Streptomyces* genes coding for synthesis of different polyketides used to clone antibiotic biosynthetic genes

F. Malpartida*||, S. E. Hallam*, H. M. Kieser*, H. Motamedi†, C. R. Hutchinson†, M. J. Butler‡, D. A. Sugden‡, M. Warren‡, C. McKillop§, C. R. Bailey§, G. O. Humphreys§ & D. A. Hopwood*

* John Innes Institute, Norwich NR4 7UH, UK

† School of Pharmacy, University of Wisconsin, Madison, Wisconsin 53706, USA

‡ Pfizer Ltd, Sandwich, Kent CT13 9NJ, UK

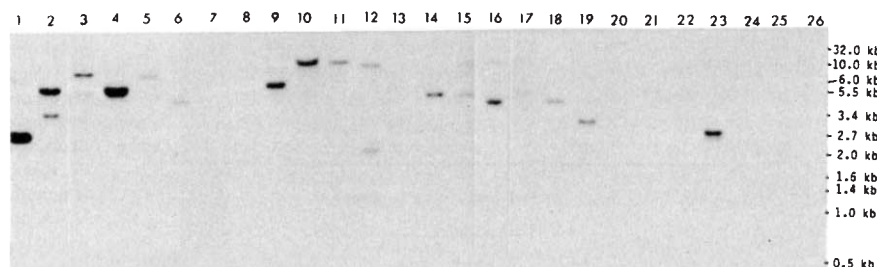
§ Apcel Ltd, 545 Ipswich Road, Slough SL1 4EQ, UK

Many important antibiotics such as tetracyclines, erythromycin, adriamycin, monensin, rifamycin and avermectins are polyketides. In their biosynthesis, multifunctional synthases¹ catalyse iterated condensation of thio-esters derived from acetate, propionate or butyrate to yield aliphatic chains of varying length and carrying different alkyl substituents. Subsequent modifications, including aromatic or macrolide ring closure or specific methylations or glycosylations, generate further chemical diversity. It has been suggested that, if different polyketide synthases had a common evolutionary origin, cloned DNA coding for one synthase might be used as a hybridization probe for the isolation of others². We show here that this is indeed possible. Study of a range of such synthase genes and their products should help to elucidate what determines the choice and order of condensation of different residues in polyketide assembly, and might yield, by *in vitro* recombination or mutagenesis, synthase genes capable of producing novel antibiotics³. Moreover, because genes for entire antibiotic pathways are usually clustered in *Streptomyces*⁴, cloned polyketide synthase genes are valuable in giving access to groups of linked biosynthetic genes.

The complete cluster of genes for actinorhodin biosynthesis was recently cloned from *Streptomyces coelicolor*². By complementation tests⁵ eight genes, identified using blocked mutants classified on co-synthetic and chemical criteria^{6,7}, were located on the cloned DNA. Class I and III mutants are blocked at the earliest steps in actinorhodin biosynthesis, because they convert intermediates secreted by mutants of all other co-synthesizing

|| Present address: Instituto de Investigaciones Biomedicas, CSIC, Facultad de Medicina, c/Arzobispo Morcillo 4, Madrid 28029, Spain.

Fig. 1 Southern hybridization of an *actI* probe to restriction digests (*Bgl*II for samples 7, 13, 14 and 24 and *Bam*HI for the other samples) of genomic DNA from various antibiotic-producing streptomycetes. Lane, organisms and antibiotic as follows: 1, *S. coelicolor* A3(2) (actinorhodin); 2, *S. violaceoruber* Tü22 (granaticin); 3, *S. sp.* AM-7161 (medermycin); 4, *S. tanashiensis* NRRL 3215 (kalafungin); 5, *S. glaucescens* ETH 22794 (tetracenomycin); 6, *S. peucetius* var. *caesius* IMRU3920 (adriamycin); 7, *S. rimosus* (Pfizer) (oxytetracycline); 8, *S. erythraeus* NRRL2338 (erythromycin); 9, *S. ambofaciens* ATCC15154 (spiramycin); 10, *S. fradiae* (Lilly) (tylosin); 11, *S. venezuelae* ATCC15068 (pikromycin, methymycin); 12, *S. antibioticus* ATCC 11891 (oleandomycin); 13, *S. griseus* IMRU3570 (candicidin); 14, *S. griseus* ETH A7796 (nonactin); 15, *S. albus* ATCC21838 (salinomycin); 16, *S. hygroscopicus* spp. *aureolacrimosus* (milbemycin); 17, *S. lasaliensis* NRRL3382R (lasalocid); 18, *S. cinnamomensis* (Lilly) (monensin); 19, *S. parvulus* ATCC12434 (actinomycin); 20, *S. curacoi* DSM40107 (curamycin); 21, *S. antibioticus* IMRU3720 (actinomycin); 22, *S. fradiae* NRRL B3357 (phosphomycin); 23, *S. venezuelae* UC2374 (chloramphenicol); 24, *S. griseus* ATCC10137 (streptomycin); 25, *S. albus* P; 26, *S. albus* G (salgiomycin). Washing was in 4 changes of 0.2×SSC, 0.1% SDS, 70 °C. (The second, weak, band of ~6 kb in the 'control' hybridization with *S. coelicolor* DNA in lane 1 represents a fragment of the DNA found to give rise to a brown pigment when introduced into *S. lividans*¹² and is distinct from the *act* DNA.) The *actI* probe was pIJ2345, which is pBR329¹³ carrying the *Bam*HI fragment between sites 14–17 (Ref. 5). (In similar experiments with *actIII* the probe was pIJ2346, which is pBR329 carrying the *Bam*HI fragment between sites 13–14.)



Genomic DNA

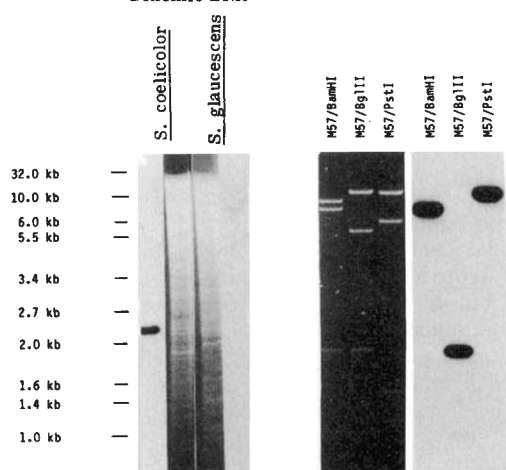
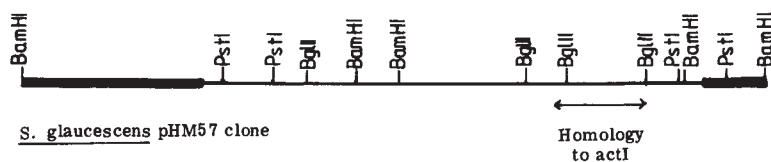


Fig. 2 Above, Southern hybridization of an *actI* probe (see Fig. 1 legend) with (left) *Bam*HI-digested genomic DNA of *S. coelicolor* and *S. glaucescens* and (right) digests of pHM57⁸ carrying *S. glaucescens* tetracenomycin biosynthetic genes cloned in pIJ702¹⁴. Below, restriction map of pHM57 showing the limits of homology with the *actI* probe deduced from the experiment. Heavy lines represent pIJ702. (Note that pHM57 carries the early genes for tetracenomycin biosynthesis; pHM152, which partially overlaps pHM57 and carries the late genes⁸, showed no hybridization with the *act* DNA.)

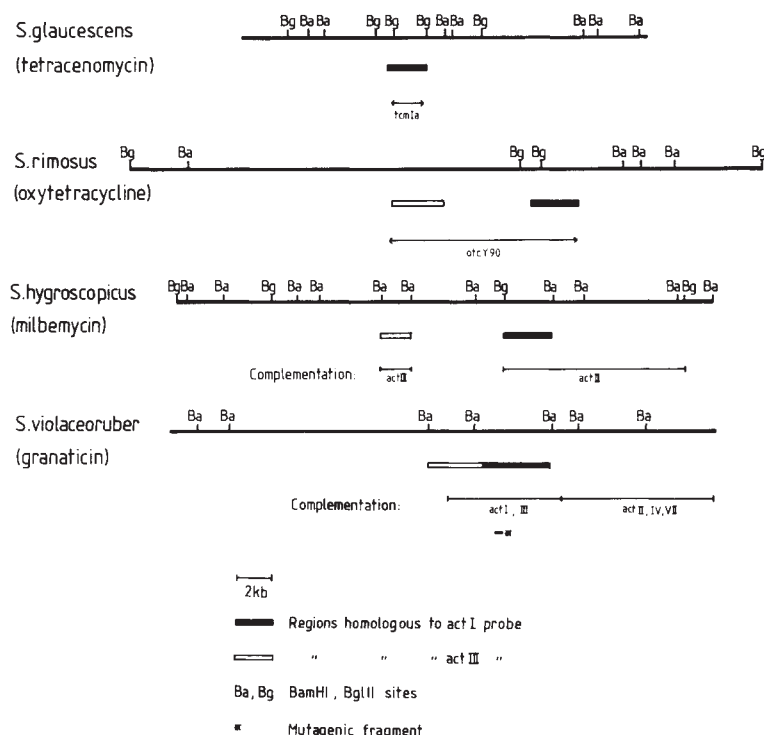


classes to actinorhodin but secrete no material capable of conversion by other mutants. Class VII mutants, blocked at the next identified step of the pathway, accumulate a shunt product derived from a complete polyketide chain⁷; thus gene VII acts after polyketide chain assembly. The amino-acid sequence of the *actIII* gene product (derived from the DNA sequence) is similar to those of ribitol dehydrogenase from *Klebsiella aerogenes* and some other dehydrogenases/reductases (S.E.H., unpublished data). This is significant in the light of the finding that 6-methylsalicylic acid synthase from *Penicillium patulum* is a complex that includes a NADPH-linked reductase acting at a specific step in polyketide assembly¹. Thus the *actI* and *actIII* gene products appear to be components of the actinorhodin polyketide synthase: *actIII* would code for a reductase and *actI* probably for the condensing enzyme itself.

DNA probes for *actI* and *actIII* were hybridized at high stringency to restriction digests of genomic DNA from *S. coelicolor* (as control) and 25 other streptomycetes. With the *actI* probe, hybridizing bands of varying intensity were seen (Fig. 1) in the DNA from 14 of 18 known polyketide-producers, including those making anthracyclines (tetracenomycin, adriamycin), macrolides (spiramycin, tylosin, pikromycin, oleandomycin), polyethers (nonactin, salinomycin, lasalocid,

monensin), and milbemycin as well as the isochromanquinone class (granaticin, medermycin, kalafungin) to which actinorhodin belongs (particularly strong hybridization was seen with two of these); only strains producing oxytetracycline, erythromycin, candicidin and curamycin showed no bands. Of the seven strains not known to produce polyketides, a chloramphenicol-producing *S. venezuelae* and actinomycin-producing *S. parvulus* and *S. antibioticus* strains showed hybridization (Fig. 1). With the *actIII* probe, hybridizing bands were seen with the same set of strains as with the *actI* probe, except that in the tetracenomycin-producing *S. glaucescens* no band hybridizing with *actIII* was revealed.

These results demonstrate a correlation, albeit incomplete, between polyketide production and the presence of DNA sequences partially homologous with the actinorhodin polyketide synthase probes. To see if some of the homologous sequences indeed represented polyketide synthase genes, previously cloned DNA for tetracenomycin C and oxytetracycline biosynthesis^{8,9} was probed with *actI* and *actIII*. Southern hybridization between the *actI* probe and clones carrying the entire set of tetracenomycin C genes (*tcm*)⁸ revealed a segment of ~2 kilobases (kb) of homologous *tcm* DNA (Fig. 2), accounting for the hybridization to *actI* in genomic digests of *S. glaucescens*



(Fig. 1). This region of the cloned *S. glaucescens* DNA complements *tcm* mutants of class Ia, which probably lie in the tetragenomycin polyketide synthase gene⁸. A reciprocal experiment in which the *tcm* DNA was used as a probe against *act* DNA showed homology only to the *actI* region. Interestingly, there was no homology between the *actIII* probe and the *tcm* gene cluster. We return to this point later.

Similar experiments identified regions with weak homology to either *actI* or *actIII* in cloned *S. rimosus* DNA⁹ carrying genes (*otc*) for oxytetracycline biosynthesis (Fig. 3). DNA from the homologous region complements *otcY90*, a mutation blocking oxytetracycline biosynthesis at one of the earliest steps¹⁰. Homology between the *act* probes and the cloned *otc* DNA was much weaker than with the *tcm* DNA, correlating with the failure to observe hybridizing bands in genomic digests of *S. rimosus* (Fig. 1).

Next we attempted to isolate genes for polyketide synthesis from the milbemycin-producing *S. hygroscopicus* ssp. *aureolacrimosus* and the granaticin-producing *S. violaceoruber* Tü22 by probing clone libraries with *actI* and *actIII* DNA. In both species, hybridizing clones were found. Restriction mapping and further hybridizations identified linked sequences homologous to *actI* and *actIII* (Fig. 3) on restriction fragments corresponding to those seen in genomic digests (Fig. 1). Functional homology of the cloned genes with those in the actinorhodin pathway of *S. coelicolor* was demonstrated by introducing the cloned *S. hygroscopicus* or *S. violaceoruber* DNA homologous to *actIII* into an *actIII* mutant of *S. coelicolor*, whereupon pigments similar or identical to actinorhodin were produced. Other classes of *act* mutants were complemented by different segments of the cloned *S. hygroscopicus* DNA (*actII*) or *S. violaceoruber* DNA (*actI*, II, IV, VII), indicating that groups of genes from polyketide biosynthetic pathways had indeed been cloned (Fig. 3). To confirm that these genes were involved in the granaticin or milbemycin biosynthetic pathways, the technique of mutational cloning was employed. In this approach mutations are generated when an *att*-site-deleted ϕ C31 phage vector forms a lysogen by integration into the host genome through recombination between a cloned DNA segment internal to a transcription unit and homologous DNA in the recipient¹¹. Granaticin non-producing or milbemycin non-producing lysogens were obtained with the test clones, indicating

Fig. 3 Restriction maps of cloned DNA involved in polyketide synthesis from four different streptomycetes. The *S. glaucescens tcm* DNA and *S. rimosus otc* DNA segments were cloned and mapped previously^{8,9}. *S. hygroscopicus* DNA was isolated in the present work by probing a colony library in *Escherichia coli* cosmid pIJ680 (T. Kieser, personal communication) with a purified *actIII* fragment (*Bam* HI sites 13–14: ref. 5); *S. violaceoruber* DNA was isolated by probing a λ EMBL4¹⁵ plaque library with pIJ2338, which is pBR329 carrying *actI*+*actIII* DNA (*Bam* HI sites 13–17: ref. 5). Regions of the four DNA segments homologous with *actI* and *actIII* from *S. coelicolor* are indicated. The abilities of certain DNA segments to complement blocked mutants of the same strain (*tcmIa* or *otcY90*) or various classes of *act* mutants of *S. coelicolor*, when introduced on pIJ941 or pIJ943¹⁶, are also shown. The bar marked with an asterisk shows the fragment of *S. violaceoruber* DNA capable of generating granaticin non-producing lysogens when cloned into ϕ C31 KC516¹⁷, thus demonstrating that the cloned DNA is part of a set of biosynthetic genes for granaticin. For *S. hygroscopicus* the 1.6-kb *Bam* HI fragment homologous to *actIII*, when cloned into ϕ C31 derivative pPOD9 (C.R.B., unpublished), gave rise to defective lysogens which failed to release phages. The lysogens were milbemycin non-producing and Southern hybridization showed them to have deletions of DNA from the chromosomal region surrounding the cloned *Bam* HI fragment. This suggests that the cloned DNA contains genes involved in milbemycin biosynthesis.

that they carried genes involved in the biosynthesis of these antibiotics (see Fig. 3 legend).

It is significant that DNA homologous to both *actI* and *actIII* is found in the oxytetracycline, granaticin and milbemycin gene clusters, but only to *actI* in the tetragenomycin DNA (Fig. 3). Unlike the other four polyketides, tetragenomycin requires no reductive step associated with carbon chain assembly. If *actIII* is indeed a reductase gene (see above), homology to *actIII* would therefore not be expected in the tetragenomycin gene cluster.

The carbon chains of four of the antibiotics—actinorhodin, granaticin, tetragenomycin and oxytetracycline—are derived by condensation of acetate-derived residues, giving chains of varying length; for milbemycin, acetate- and propionate-derived residues must be condensed in the correct sequence. Moreover, as shown in Fig. 1, there is circumstantial evidence for homology between the *act* probes and DNA involved in the biosynthesis of other classes of antibiotics, including macrolides and polyethers with further patterns of chain assembly. Thus the cloning strategy demonstrated here is likely to be widely applicable. It might be extended by the use of probes to recognize cryptic producers of secondary metabolites of a particular class. Thus, *S. parvulus*, one of the apparent polyketide non-producers that contained DNA sequences hybridizing with the *actI* probe (Fig. 1), was recently found to have a normally unexpressed capacity for polyketide biosynthesis, producing the polyether nonactin when its ability to make the non-polyketide actinomycin was blocked by mutation (U. Keller, personal communication). The principles established here for polyketides could probably be extended to other chemical classes of antibiotics such as β -lactams, aminoglycosides or glycopeptides.

We thank Evelyn Wendt-Pienkowski for technical assistance. This work was supported by grants to the John Innes Institute from the AFRC of the UK and the John Innes Foundation, to the University of Wisconsin from the NIH (CA35381), and to Apel from Agrigenetics Research Associates Ltd and Lubrizol Genetics Inc.

Received 3 November 1986; accepted 6 January 1987.

1. Dimroth, P., Walter, H. & Lynen, F. *Eur. J. Biochem.* **13**, 98–110 (1970).
2. Malpartida, F. & Hopwood, D. A. *Nature* **309**, 462–464 (1984).
3. Hopwood, D. A. *et al. Nature* **314**, 642–644 (1985).

4. Hopwood, D. A. in *Biological, Biochemical and Biomedical Aspects of Actinomycetes* (eds Szabo, G., Biro, S. & Goodfellow, M.) 3-14 (Akademiai Kiado, Budapest, 1986).
5. Malpartida, F. & Hopwood, D. A. *Molec. gen. Genet.* **205**, 66-73 (1986).
6. Rudd, B. A. M. & Hopwood, D. A. *J. gen. Microbiol.* **114**, 35-43 (1979).
7. Floss, H. G. et al. in *Regulation of Secondary Metabolite Formation* (eds Kleinkauf, H., von Döhren, H., Dormauer, H. & Neemann, G.) 283-304 (VCH, Weinheim, 1986).
8. Motamedi, H., Wendt-Pienkowski, E. & Hutchinson, C. R. *Abstr. Fifth Int. Symp. Genet. ind. Microorg. Split, Yugoslavia*, 159 (1986).
9. Butler, M. J., Hunter, I. S., Sugden, D. A. & Warren, M. *Abstr. Fifth Int. Symp. Genet. ind. Microorg. Split, Yugoslavia*, 161 (1986).
10. Rhodes, P. M., Winskill, N., Friend, E. J. & Warren, M. *J. gen. Microbiol.* **124**, 329-338 (1981).
11. Chater, K. F. & Bruton, C. J. *Gene* **26**, 67-78 (1983).
12. Horinouchi, S. & Beppu, T. *J. Bact.* **162**, 406-412 (1985).
13. Covarrubias, L. & Bolivar, F. *Gene* **17**, 79-89 (1982).
14. Katz, E., Thompson, C. J. & Hopwood, D. A. *J. gen. Microbiol.* **129**, 2703-2714 (1983).
15. Frischhauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. *J. molec. Biol.* **170**, 827-842 (1983).
16. Hopwood, D. A. et al. *Genetic Manipulation of Streptomyces, a Laboratory Manual* (John Innes Foundation, Norwich, 1985).
17. Rodicio, M. R., Bruton, C. J. & Chater, K. F. *Gene* **34**, 283-292 (1985).

Poly(dA)·poly(dT) is a B-type double helix with a distinctively narrow minor groove

D. G. Alexeev, A. A. Lipanov & I. Ya. Skuratovskii

Institute of Molecular Genetics, USSR Academy of Sciences,
46 Kurchatov Square, Moscow 123182, USSR

The structure of poly(dA)·poly(dT) currently arouses great interest, mainly because dA_n·dT_n stretches are associated with considerable DNA bending^{1,2}. Until recently the heteronomous DNA described by Arnott *et al.*³, with the poly(dA) and poly(dT) chains in A and B conformations respectively, was the only detailed model of this structure. Following our earlier studies of the interaction of DNA and monovalent ions⁴⁻⁶, we examined the X-ray diffraction of the bivalent Ca²⁺ salt of poly(dA)·poly(dT) (Ca-poly(dA)·poly(dT)) and found no sign of a heteronomous structure: Ca-poly(dA)·poly(dT) in fibres shows fully equivalent B-type conformations of the opposite sugar-phosphate chains. A revision of the structure of the sodium salt, Na-poly(dA)·poly(dT), based on this result, yields only a slightly heteronomous structure with each chain in a B-type conformation, which is in much better agreement with the experimental data underlying the original heteronomous model³. Both structures, Ca- and Na-poly(dA)·poly(dT), have a minor groove narrower than that of the B form: this peculiarity seems to be very important for the interaction of poly(dA)·poly(dT) and biologically significant molecules (including proteins and antibiotics). The specific base-pair positions in poly(dA)·poly(dT) may account for the DNA bending adjacent to dA_n·dT_n tracts.

Figure 1 shows the diffraction pattern of Ca-poly(dA)·poly(dT). Some important features of the crystal and molecular structure were revealed by comprehensive Patterson analysis.

Figure 2a and b shows cylindrically averaged Patterson functions for Ca-poly(dA)·poly(dT) and for Li-B-DNA, for which a more reliable structure has been obtained^{3,7}. These functions point to the similarity of the molecular structures of Li-B-DNA and Ca-poly(dA)·poly(dT) and to the fact that in the Ca-poly(dA)·poly(dT) unit cell, as in the classic crystalline B form, two molecules are related by translation. The systematic reflection absences in the *hk0* plane and the Patterson projection for the *0kl* plane also point to translation. We may conclude that, the T-A and A-T base pairs being indistinguishable (at the existing resolution of about 0.3 nm), the presence of the translation relation indicates a dyadic relation between the conformations of the poly(dA) and poly(dT) chains of a molecule (see Fig. 2 legend). Poly(dG)·poly(dC) in the A form⁸ is a most convincing example of such a situation, where the molecular dyad is also a crystallographic symmetry element.

As a result the heteronomous model³ proves unacceptable for Ca-poly(dA)·poly(dT), for it must lead to the disappearance of

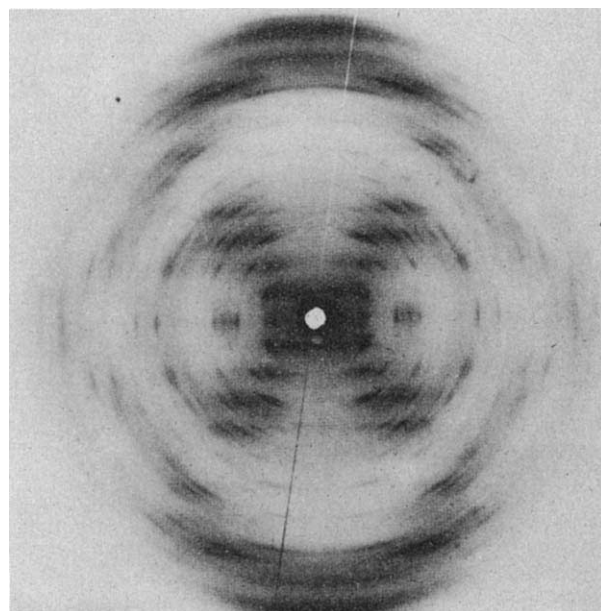


Fig. 1 Diffraction pattern of Ca-poly(dA)·poly(dT) (81% r.h.) in which the intensities of 87 independent diffraction spots were measured. Analysis of this pattern shows an orthorhombic lattice; the unit cell ($a = 1.871$ nm, $b = 4.032$ nm, $c = 3.232$ nm, $\alpha = \beta = \gamma = 90^\circ$) contains two polynucleotide molecules. As in the case of the sodium salt³, the pitch of the tenfold double helix is 3.23 nm and the molecules are packed into the same kind of flat layers in the *a* direction, although in our case the distance between them (in the *b* direction) is 0.48 nm larger. The structure was refined by the linked-atom technique (LALS program⁹), considerably modified in our case. We used a one-parameter flexible sugar ring based on the statistical analysis of the nucleoside and nucleotide structures from the Cambridge database³⁰. Compared to ref. 3, we did not introduce any additional parameters defining the position of bases. The *R* value ($R = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$) for the refined structure of Ca-poly(dA)·poly(dT) is 0.31, and $R'' = (\sum (F_{\text{obs}} - F_{\text{calc}})^2 / \sum F_{\text{obs}}^2)^{1/2} = 0.36$. R'' was calculated taking into account unobserved reflections in those cases where the calculated amplitudes exceeded the estimated threshold. Water correction has been made according to ref. 7.

translation peaks corresponding to the position of the second molecule in the unit cell (Fig. 2a). Then the only possible space group (as in the case of Li-B-DNA) is $P2_12_12_1$. A three-dimensional Patterson synthesis independently confirms the presence of translation and provides an image of the polynucleotide molecule (Fig. 2c). As a result of this procedure we have obtained the azimuth orientation of the molecule in the unit cell and a fairly good approximation of the phosphate group positions.

The sugar conformations in the refined Ca-poly(dA)·poly(dT) structure are close to C2'-endo. The search for models with different conformations for poly(dA) and poly(dT) chains has shown the Ca-poly(dA)·poly(dT) structure to lack heteronomy within the accuracy of the linked-atom technique (LALS method)⁹.

Because the packing and molecular parameters of Na-poly(dA)·poly(dT) are close to those of the calcium salt, we attempted to adapt our model of Ca-poly(dA)·poly(dT) to the X-ray diffraction data from ref. 3. The result is a Na-poly(dA)·poly(dT) structure in which the poly(dA) and poly(dT) chain conformations differ much less than in the heteronomous model and the whole structure is fairly close to Ca-poly(dA)·poly(dT). The newly obtained structure of Na-poly(dA)·poly(dT) is free of overshoot contacts of non-bonded atoms (intramolecular or intermolecular) and has an *R* value of 0.26 ($R'' = 0.29$). Under the same calculation conditions the

Fig. 2 *a, b*, Cylindrically averaged Patterson functions for (a) Ca-poly(dA)·poly(dT) and (b) Li-B-DNA. The vertical size of the figures corresponds to a half-length of the *c* side of the unit cell: 1.616 nm for *a* and 1.690 nm for *b*. The peak of the function with coordinates $r=0$ and $z=0$ (number 1) corresponds to zero interatomic vectors. The DNA double helix diameter is ~ 2 nm, as intramolecular peaks are located in the $r < 2$ nm region and only intermolecular peaks are located in the $r > 2$ nm region. In the case of Li-B-DNA⁷ we know the positions of the two molecules in the unit cell, so that we can find the Patterson function peaks corresponding to all intermolecular translations. Thus, peaks 3 and 6 correspond to lattice translations (along sides *b* and *a*⁷). Peaks 2 and 4 correspond to non-lattice translations relating the two molecules in the unit cell, which are shifted relative to each other by $\sim c/3$ (ref. 7) along the *z* axis. The nucleotide sequence is random in the case of Li-B-DNA, which makes such translations obvious, but in the case of Ca-poly(dA)·poly(dT) there is also a complete set of intermolecular translation peaks similar to those of Li-B-DNA. As it is highly unlikely that the long poly(dA)·poly(dT) molecules would form crystallites with all molecules directed one way, the translation peaks must indicate the presence of traditional molecular dyads running perpendicular to the helix axis. Here peaks 3 and 5 correspond to lattice translations, which are equal to $|a|$ and $|2a|$, and peaks 2 and 4 correspond (as with Li-B-DNA) to non-lattice translations, the relative shift (Δz) of the two molecules in the unit cell being $\sim c/3$. *c*, Electron density square for Ca-poly(dA)·poly(dT) molecule (minus the constant), calculated according to the three-dimensional Patterson synthesis (within the framework of space group $P2_12_12_1$). The projection is made down the helix axis. The basis of this approach is similar to that of the previously described 'helical projection' derived from a cylindrically averaged Patterson function³¹. The squared electron density maxima are in good agreement with the superimposed sugar-phosphate chains of the refined model. The radii of phosphorus atoms are shown (large circles) reduced 5 times compared to their van der Waals radii; the other atomic sizes are reduced 8 times.

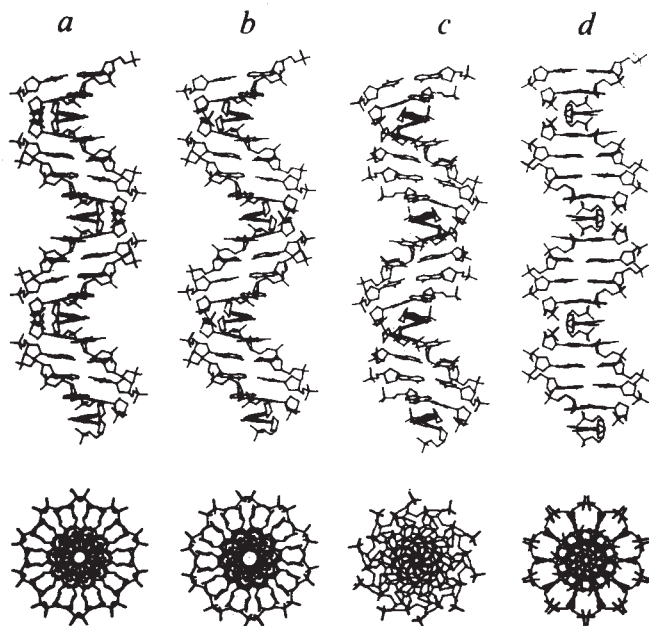
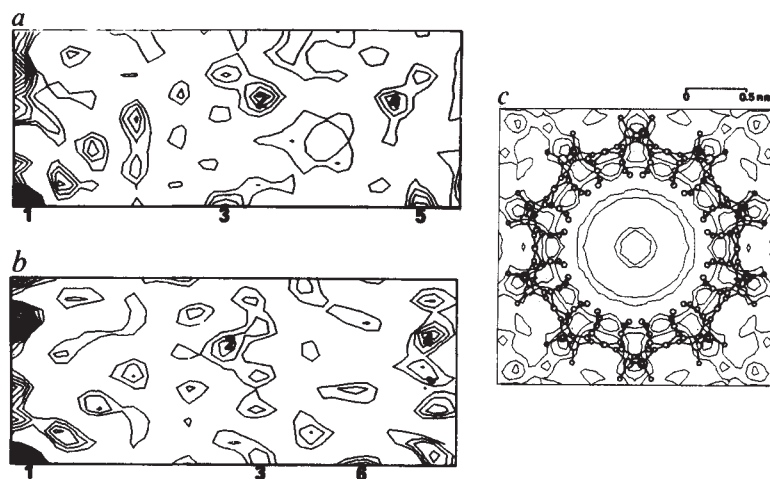


Fig. 3 Projections perpendicular to (above) and down (below) the helix axis for: *a*, the calcium salt; *b*, the sodium salt of poly(dA)·poly(dT); *c*, the heteronomous model of Na-poly(dA)·poly(dT)³; *d*, poly(dA)·poly(dT) in the classic B form³. The width of the minor groove in these structures (the shortest distance between the phosphorus atoms of opposite chains) is 0.92 nm, 0.92 nm, 1.1 nm and 1.2 nm respectively. The conformation at 03'–P is *gauche minus* in our models of poly(dA)·poly(dT) and not *trans* as in the B-DNA³. This is in agreement with the crystal data¹¹. The bases are shifted about 0.1 nm away from the helix axis (in the same direction as in A-DNA). The shift and tilt of -6° account for the reduced pitch in the poly(dA)·poly(dT) double helix (by 0.15 nm) compared with the B form⁷. The standard deviation for the backbone torsion angles between poly(dA) and poly(dT) chains is 10° in our Na-poly(dA)·poly(dT) model, whereas in the heteronomous model³ it is 45° . In our case the effective relative shift (Δz) of the two counter-directed Na-poly(dA)·poly(dT) molecules in the unit cell (space group $P2_1$) is close to zero, and the azimuth orientation of the molecules is close to that in Ca-poly(dA)·poly(dT), where one of the pseudodyads is directed along the *b* side. The situation is essentially different in the packing of heteronomous molecules (S. Arnott, personal communication), although in both cases the molecules are situated at the vertices of a 72° rhombus. A full set of molecular and crystal packing parameters will be given elsewhere.

heteronomous model has an *R* value of 0.28 ($R''=0.37$) even though it was developed with a greater number of parameters³, and so can be rejected with 99.5% confidence according to Hamilton's statistical significance test¹⁰. Figure 3 shows the optimized structures of Ca- and Na-poly(dA)·poly(dT). The observed base positions (the negative tilt of -6° and the large positive propeller twist of 20°) in both structures considerably narrow down the minor groove of the double helix (by 0.28 nm compared with the classic B form).

These new structures show a striking similarity to the A·T stretch of the CGCGAATTCGCG dodecamer in crystals^{11,12} (for example, the shortest distances between phosphates virtually coincide with those in Fig. 3). One of the important

features of the dodecamer is the double-layer hydration spine in the minor groove of the A·T stretch, which stabilizes the B-type structure^{12,13}. An earlier study of Cs-B-DNA fibres⁵ revealed an ionic stabilization mechanism of the B form which is similar to the water mechanism. Model estimations of the sodium ion positions in the newly found structure of Na-poly(dA)·poly(dT) have shown that only the latter mechanism is possible (see Fig. 4). A double-layer water spine should be present inside the minor groove in the case of Ca-poly(dA)·poly(dT) as well, because the Ca^{2+} ions that could, in principle, be situated there, are in reality located near the phosphate groups, as attested by the results of extended X-ray absorption fine structure (EXAFS) spectroscopy^{14,15}.

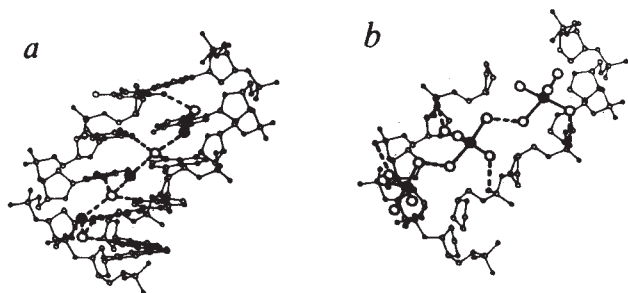


Fig. 4 *a*, The double-layer water spine in the minor groove of poly(dA)·poly(dT) similar to that observed in the CGCGAATTCGCG dodecamer^{12,13}. Five successive nucleotide pairs are shown. Dashed lines denote hydrogen bonds. Second layer water molecules are crossed. According to our previous calculations for the poly(dA)·poly(dT) molecule in the B form, Na⁺ ions, like Cs⁺ ions⁵ or water molecules of the first layer, can enter into close coordination with N3 atoms of adenines and O2 atoms of thymines (see Figure of ref. 4). In the newly determined structure of Na-poly(dA)·poly(dT), such a model is at variance with the narrowed minor groove: very strong distortions occur in the (O2/N3)-Na⁺-(H₂O)₄ octahedron. By contrast, the octahedral coordination of Na⁺ can be kept practically undistorted if the ions are positioned in the third solvent layer adjacent to the double-layer water spine. *b*, A model of the integration of Na⁺ ions into the poly(dA)·poly(dT) structure on the side of the minor groove. The base atoms and the double-layer water spine are not shown. Ions (black circles) are situated over the second-layer water molecules (see *a*) and interact with them. Optimized sodium-water octahedrons are linked by hydrogen bonds both to each other and to the phosphate oxygens of the opposite chains in the double helix. Such an ion-water spine meets with no steric obstacles in the crystal lattice.

These conclusions are supported by the energy calculations of Chuprina¹⁶⁻¹⁸, which suggest that a water spine in the minor groove of poly(dA)·poly(dT) would lead to an energetically optimal structure with the distinctive features we observe in fibres. It seems quite likely that the water spine makes a substantial contribution to the stabilization of the poly(dA)·poly(dT) structure and is largely responsible for its marked conservatism in different circumstances. The molecular structures of Ca- and Na-poly(dA)·poly(dT) are similar; they do not adopt the A form; moreover they retain the B' structure, virtually unchanged, at any relative humidity. This polynucleotide does not adopt the A form in water-alcohol solutions either¹⁹. Furthermore, even short (4 base-pair) dA_n·dT_n insertions in a quasi-random DNA sequence retain the fibre-characteristic helical periodicity of 10.0 base pairs per turn in solution^{20,21}, whereas for other sequences in solution this quantity changes to 10.6 ± 0.1²⁰⁻²².

Our results support the concept that the junctions between this conservative structure and the B form are responsible for the equilibrium bends in some natural DNA fragments^{1,23,24}, for the equilibrium base positions in our structures of Ca- and Na-poly(dA)·poly(dT) differ from their positions in the B form. The proposed direction of bending¹ is in good agreement with structural findings presented here. The observed poly(dA)·poly(dT) structure in fibres also testifies in favour of the narrowed minor groove in A-T sequences. Experiments involving the cleavage of DNA by some nucleases^{25,26}, the binding of non-intercalating drugs such as netropsin^{27,28} or the binding of some minor-groove recognition proteins²⁹ suggests that the distinctive narrowness of the minor groove may have specific biological significance.

We thank Professor M. D. Frank-Kamenetskii for valuable discussions, Professor S. Arnott for the X-ray structure amplitudes for Li-B-DNA, Dr V. Chuprina for making his data available before publication and Mr D. Agrachev for his help with the manuscript.

Received 7 July; accepted 30 December 1986.

1. Koo, H.-S., Wu, H.-M. & Crothers, D. M. *Nature* **320**, 501-506 (1986).
2. Hagerman, P. J. *Nature* **321**, 449-450 (1986).
3. Arnott, S. *et al. Nucleic Acids Res.* **11**, 4141-4145 (1983).
4. Bartenev, V. N., Golovanov, E. I. & Skuratovskii, I. Ya. in *Symposium on Biophysics of Nucleic Acids and Nucleoproteins*, Tallinn, 179 (Academy of Sciences of the Estonian SSR, Tallinn, 1981).
5. Bartenev, V. N. *et al. J. molec. Biol.* **169**, 217-234 (1983).
6. Lipanov, A. A. *et al. Biofizika* **31**, 336-338 (1986).
7. Langridge, R. *et al. J. molec. Biol.* **2**, 38-64 (1960).
8. Arnott, S. & Selsing, E. *J. molec. Biol.* **88**, 551-552 (1974).
9. Smith, P. J. C. & Arnott, S. *Acta crystallogr.* **A34**, 3-11 (1978).
10. Hamilton, W. C. *Acta crystallogr.* **18**, 502-510 (1965).
11. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. *J. biol. Chem.* **257**, 14686-14707 (1982).
12. Dickerson, R. E., Kopka, M. L. & Pjura, P. in *Biological Macromolecules and Assemblies* Vol. 2 (eds McPherson, A. & Jurnac, F.) (Wiley, New York, 1985).
13. Kopka, M. L., Fratini, A. V., Drew, H. R. & Dickerson, R. E. *J. molec. Biol.* **163**, 129-146 (1983).
14. Alexeev, D. G. *et al. in Daresbury Annual Report 1983/84 Appendix*, 57 (SERC Daresbury Laboratory, Warrington, 1984).
15. Skuratovskii, I. Ya., Hasnain, S. S., Alexeev, D. G., Diakun, G. P. & Volkova, L. I. *J. inorg. Biochem.* (in the press).
16. Chuprina, V. P. *FEBS Lett.* **186**, 98-102 (1985).
17. Chuprina, V. P. *FEBS Lett.* **195**, 363-364 (1986).
18. Chuprina, V. P. *Nucleic Acids Res.* (in the press).
19. Ivanov, V. I., Minchenkova, L. E., Minyat, E. E. & Schyolkina, A. K. *Cold Spring Harb. Symp. quant. Biol.* **47**, 243-250 (1983).
20. Peck, L. J. & Wang, J. C. *Nature* **292**, 375-378 (1981).
21. Rodes, D. & Klug, A. *Nature* **292**, 378-380 (1981).
22. Horowitz, D. S. & Wang, J. C. *J. molec. Biol.* **173**, 75-91 (1984).
23. Wu, H.-M. & Crothers, D. M. *Nature* **308**, 509-513 (1984).
24. Diekmann, S. & Wang, J. C. *J. molec. Biol.* **186**, 1-11 (1985).
25. Drew, H. R. & Travers, A. A. *Cell* **37**, 491-502 (1984).
26. Drew, H. R. & Travers, A. A. *J. molec. Biol.* **186**, 773-790 (1985).
27. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. *J. molec. Biol.* **183**, 553-563 (1985).
28. Dickerson, R. E. & Kopka, M. L. *J. biomolec. Struct. Dyn.* **3**, 423-431 (1985).
29. Solomon, M. J., Strauss, F. & Varshavsky, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1276-1280 (1986).
30. Bartenev, V. N., Kameneva, N. G. & Lipanov, A. A. *Acta crystallogr.* (in the press).
31. Klug, A., Crick, F. H. C. & Wyckoff, H. W. *Acta crystallogr.* **11**, 199-213 (1958).

Long-range cooperativity between gene regulatory sequences in a prokaryote

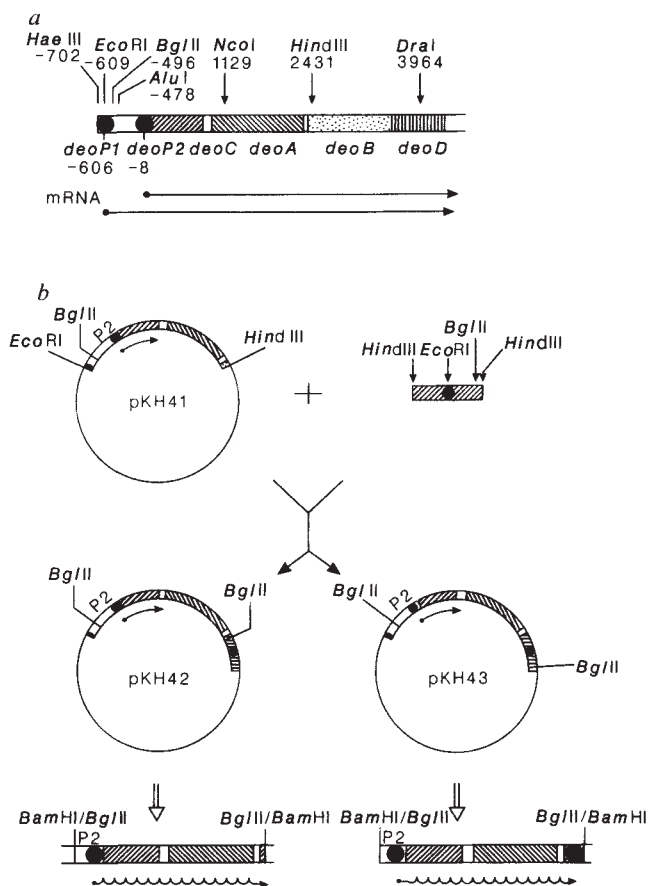
Gert Dandanell*, Poul Valentin-Hansen†, Jens Erik Løve Larsen† & Karin Hammer*

* Institute of Biological Chemistry B, University of Copenhagen, Sølvgade 83, DK-1307 Copenhagen, Denmark
† Department of Molecular Biology, University of Odense, Campusvej 55, DK-5230 Odense M, Denmark

Regulation of transcription initiation by proteins binding at DNA sequences some distance from the promoter region itself seems to be a general phenomenon in both eukaryotes and prokaryotes. Proteins bound to an enhancer site in eukaryotes can turn on a distant gene, whereas efficient repression of some prokaryotic genes such as the *gal*, *ara* and *deo* operons of *Escherichia coli*, requires the presence of two operator sites, separated by 110, 200 and 600 base pairs (bp) respectively¹⁻⁶. In the *deo* operon, which encodes nucleoside catabolizing enzymes, we have shown that efficient and cooperative repression can be obtained when the distance between the two sites ranges from 224 to 997 bp^{5,6}. Here, we report that transcription initiation can be regulated from an operator site placed 1 to 5 kilobases (kb) downstream of the *deoP2* promoter (and downstream of the transcribed gene), and present the first experimental data for prokaryotic regulation at distances >1 kb. Our results support the model of DNA loop formation as a common regulatory mechanism explaining both some prokaryotic regulation and the action of eukaryotic enhancers⁷.

Full *deoR* repression of the *deoP1* and *deoP2* promoters requires the presence of three *deoR* binding sites^{5,6}. Two of these operators, 599 bp apart, overlap the sites of transcription initiation (*deoP1* and *deoP2*, Fig. 1a) whereas the third site O_E is located 279 bp upstream of *deoP1*^{5,6,8}. The *deoP1* and *deoP2* operator sites alone are sufficient to obtain a *deoR* regulation

Fig. 1 *a*, Structure of the *deo* operon. The *deo* genes code for deoxy-riboaldolase (*deoC*), thymidine phosphorylase (*deoA*), phosphopentomutase (*deoB*) and purine nucleoside phosphorylase (*deoD*). Tetracistronic transcripts are initiated from two promoters *deoP1* and *deoP2* which each contain a *deoR* binding site (black circle) overlapping the -10 region of the promoter (refs 5, 8; ref. 12 for review). The third *deoR* binding site *O_E* located 279 bp upstream of *deoP1*⁶, is not shown here as it is not present in any of the plasmids used in this work. Restriction sites are indicated and numbered relative to the start of mRNA from *deoP2*. *b*, Construction of plasmids used in Fig. 3. The vector used is the mini R1 plasmid pJEL126. This 19.5-kb plasmid is present at only one copy per genome, and contains unique *EcoRI* and *BamHI* sites proximal to *lacZ*^{6,9}. A 3,040-bp *EcoRI*-*HindIII* fragment containing the *deoP2*-*deoC*-*deoA* operon was isolated from pVH19 (P.V.-H., in preparation) and inserted into a pBR322 derivative (pGD3) resulting in pKH41. This plasmid contains unique *BglII* and *HindIII* sites on each side of the *deoP2*-*deoC*-*deoA* operon. A 224-bp *HaeIII*-*AluI* fragment containing the *deoP1* operator⁵ was inserted into the *HindIII* site of pKH41 after ligation of a *HindIII* linker (pKH42 and pKH43). By transferring the *BglII* fragment from these plasmids into the *BamHI* site of the low-copy-number plasmid pJEL126⁶, two plasmids were constructed: pGD108 contains the *deoP2* promoter/operator followed by the *deoC* and *deoA* genes and the *deoP1* operator placed 2,540 bp downstream of the *deoP2* operator. The other plasmid pGD103 contains only the *deoP2* operator. Plasmids pKH162, pKH163 and pKH165 (see Fig. 3) were derived from pKH41 by a similar procedure, pKH41 first being digested with *NcoI*, the sticky ends filled in with DNA polymerase (Klenow fragment) and a 12-bp *HindIII* linker ligated to the blunt ends (pGD35). This new *HindIII* site was used for insertion of the *deoP1* operator and the entire construction was transferred to pJEL126 on a *BglII* fragment as above. In pKH165 a *deoP1* operator fragment was used in which the operator had been destroyed by filling in the *EcoRI* site in the centre of the *deoR* binding site⁵. To construct pGD110 and pGD107 the *BglII* fragment of pKH42 and pKH43 was first inserted into pGD201, a pBR322 derivative containing a unique *BglII* site and no *HindIII* sites. In the resulting plasmids pGD203 and pGD204 the *HindIII* site (position 2,431) of the *deo* operon is now unique and the entire *deo* operon was thereafter reconstructed by inserting a *HindIII*-*DraI* fragment (2,431-3,964). This DNA fragment was isolated from pVH7 (P.V.-H., in preparation) and a *HindIII* linker was ligated to the *DraI* end. The *BglII* fragment of each of the resulting plasmids (pGD205 and pGD206) was inserted into the *BamHI* site of pJEL126 as described above giving pGD110 and pGD107 respectively.



of 20-30-fold compared with 2-fold when only one of the promoter/operator sites is present^{5,6}. In this study we asked first whether it is possible to obtain regulation of transcription initiation from an operator site positioned downstream of the transcribed region, and second, if so, to what extent repression depends on the distance between the two operator sites.

First, we used low-copy-number plasmids encoding *deoP2* gene fusions between *deoC* or *deoA* and *lacZ* (refs 6 and 9; Fig. 2, lines *a* and *d*). When the *lacZ* fusion is transcribed from *deoP2* containing the *deoP2* operator site alone, the *deoR* regulation is 1.8-fold (lines *a* and *d*). When a second *deoR* binding site is placed 567 bp upstream of the *deoP2* operator a drastic decrease in the wild-type level is observed resulting in a 23-fold regulation by *deoR* (line *b*). Next we inserted the second operator site downstream of the *lacZ* gene in the *deo-lacZ* fusions, which resulted in a small but significant increase in the *deoR* regulation (lines *c* and *e* compared with lines *a* and *d*). The distance between the operator sites in these two constructions is 3,702 and 4,685 bp and the *deoR* regulation is 3.6- and 3.0-fold respectively.

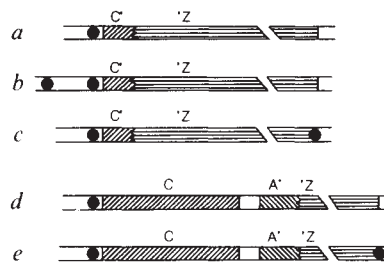
To determine in more detail how repression depends on the distance between the two operator sites, we used low-copy number vectors containing the *deoP2* promoter and varying parts of the *deo* operon (Figs 1 and 3). Expression of the first structural gene, *deoC*, from these plasmids was determined in both wild-type and *deoR* cells (Fig. 3). In accordance with the results in Fig. 2, transcription initiation from *deoP2* is regulated weakly by the *deoR* repressor (1.6-2-fold) when only one *deoR* binding site is present on the plasmid. Insertion of a second *deoR* binding site downstream of the *deoP2* operator increases

the *deoR* repression in all three experiments shown in Fig. 3. The distance between the operator sites was 1,245, 2,540 and 4,076 bp and repression was 21-, 8.8- and 5.2-fold, respectively. This inverse relationship between the extent of repression and the distance separating the two operator sites is evident when the data from Fig. 3 are used in a double-log plot, where a straight line is observed (data not shown). Extrapolation of this line to <1 kb is probably not meaningful because the resistance of the DNA molecule to bending and twisting over such short distances is high^{10,11}. Extrapolation towards long distances indicates that a twofold repression should be obtained in these fusions when the two operators are separated by 8-10 kb. A twofold *deoR* repression corresponds to that observed for *deoP2* alone, and thus, the cooperativity between the two *deoR* binding sites in our system is expected to be lost at 8-10 kb. We have tested a distance of 39 kb by using an *E. coli* strain containing two λ phages inserted in tandem each carrying a *deoP1*-*galK* fusion with only one *deoR* operator. As expected, no increase in *deoR* repression was observed (data not shown).

We have included three different controls in Fig. 3. First, insertion of a mutated *deoR* binding site downstream of the *deoP2* operator results in the same low degree of repression as that observed for the *deoP2* plasmid with no insert (pKH165 compared with pKH162). The additional repression observed when two operator sites are present on the same plasmid is therefore due to the presence of the second *deoR* binding site. Second, the twofold repression seen in the three constructions containing only one *deoR* binding site shows that the *deo* operon does not contain any *deoR* binding sites between *deoP2* and *deoD* (pKH162, pGD103 and pGD110; see also Fig. 2, lines *a*

Fig. 2 Effects of a second *deoR* operator site on *lacZ* expression from *deoP2*-*lacZ* fusions in the presence and absence of *deoR* repressor. Left, constructions as follows: *a*, pJEL190; *b*, pJEL190-O₁A; *c*, pJEL190-O₁B; *d*, pJEL194; and *e*, pJEL194-O₁B. C and A, *deoA* gene and *deoC* gene (diagonal hatching); Z, *lacZ* gene (horizontal hatching); ', side of gene deleted; ●, operator sites. Right, distance between operators and *lacZ* expression for each plasmid, in wild-type (wt) and *deoR* background. Repression is given as the ratio of expression in *deoR* to that in wild-type.

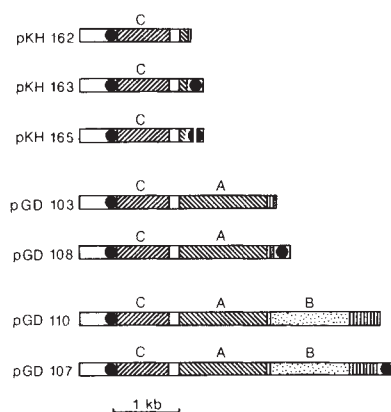
Methods. The *deo*-*lacZ* protein fusions pJEL190 (*a*) and pJEL194 (*d*)⁶ contain a unique *Asu*II restriction site in the *lacY* gene and two *Cl*aI sites, one in the *lacZ* gene and the other in *deo* DNA, 555 bp upstream of the start point of *deoP2* transcription⁸. A purified *Cl*aI fragment from pVH17⁸ containing *deo* sequences 21 bp upstream to 54 bp downstream of the central base pair in the *deoP1* operator was inserted into pJEL190 partially digested with *Cl*aI and into the unique *Asu*II site of pJEL190 and pJEL194, resulting in plasmids pJEL190-O₁A, pJEL190-O₁B and pJEL194-O₁B. Measurements of β -galactosidase expressed from fusion plasmids were done as described by Miller¹⁶. Samples for enzyme assays were taken at OD₄₅₀ in the range 0.1–0.5. Specific activities of β -galactosidase are expressed as OD₄₂₀/OD₄₅₀ ml⁻¹ min⁻¹. Cells were grown exponentially at 36 °C in AB medium¹⁷ with glycerol as carbon source. All experiments were performed with the *E. coli* K-12 strains S ϕ 928 (Δ *deo*, Δ *lac*) and S ϕ 930 (Δ *deo*, Δ *lac*, *deoR*)¹⁸.



<i>lacZ</i>			
Distance (bp)	wt	<i>deoR</i>	Repression
-	0.5	0.9	1.8
567	0.06	1.4	23.0
3702	0.22	0.8	3.6
-	0.65	1.2	1.8
4685	0.44	1.3	3.0

Fig. 3 Effects of a second operator site on *deoP2* expression of the *deo* operon. Left, plasmid constructions containing *deo* genes. C, A, B: *deoC*, *deoA*, *deoB* genes respectively; ●, operator sites; vertical hatching, parts of the *deoD* gene. Right, distance between operators and *deoC* and *deoA* expression in wild-type (wt) and *deoR* backgrounds. Repression is given as the ratio of expression in *deoR* to that in wild-type.

Methods. The second operator site was inserted into the *Nco*I, *Hind*III or *Dra*I sites as in Fig. 1. The activity of plasmid-encoded deoxyriboaldolase (*deoC*) and thymidine phosphorylase (*deoA*) was measured in extracts of S ϕ 928 and S ϕ 930 harbouring the plasmids. Cells containing the recombinant low-copy-number plasmids were grown exponentially at 33 °C for 5 generations in AB medium¹⁷ supplemented with 25 μ g ml⁻¹ ampicillin and 0.2% glycerol as carbon source. At a cell density of 1.5×10^8 , cells were collected by centrifugation, washed in $\frac{1}{2}$ vol. AB and finally resuspended in 1/60 vol. 100 mM Tris-HCl pH 7.6, 2 mM EDTA, and sonicated for 60 s. The supernatant was used for the determination of deoxyriboaldolase and thymidine phosphorylase as described^{19,20}. Units are expressed as nmol of substrate converted per min at 37 °C. The numbers given are specific activities, expressed as units of enzyme activity per mg of protein determined according to Lowry *et al.*²¹. Numbers represent the mean of at least three independent experiments. We used the low-copy plasmids where *deoP2* transcription reads in the opposite direction to that of the *lacZ* gene. Quantitatively similar results were obtained with the constructions where *deoP2* reads into *lacZ*, but in these plasmids the basal level is slightly elevated due to an unknown plasmid promoter activity (data not shown).



<i>deoC</i>				<i>deoA</i>			
Distance (bp)	wt	<i>deoR</i>	Repression	wt	<i>deoR</i>	Repression	
-	191	389	2.0	[Diagonal hatching]			
1245	17	357	21.0				
-	210	361	1.7				
-	217	422	1.9	195	340	1.7	
2540	49	427	8.8	43	282	6.6	
-	281	444	1.6	588	954	1.6	
4076	85	444	5.2	216	997	4.6	

and *d*). Third, when the activity of the *deoC* and *deoA* gene products was measured for the four plasmids carrying both genes, the *deoR* regulation was found to be the same (Fig. 3).

The main conclusion from the results in Figs 2 and 3 is that it is possible to construct an operon in *E. coli* where transcription initiation is regulated from an operator site at a position downstream of the transcribed genes. The degree of repression is inversely related to the distance between the operator sites and can be observed even at a distance of 4.7 kb.

The cooperativity between the *deoR* binding sites is the same whether they are present on a low-copy-number plasmid (mini R1)⁶ or on the bacterial chromosome (λ phage)⁵. One requirement is, however, that the binding sites are present on the same DNA molecule because the additional *deoR* binding sites carried on the bacterial chromosome at *nupG* and *tsx*¹² have no effect on the repression in our experiments. We have observed cooperativity at the following distances: 224, 279, 449, 567, 599, 603, 798, 878, 997, 1,245, 2,540, 3,702, 4,076 and 4,685 bp (see also refs 5 and 6). The DNA loop model^{2,3,5} can be used to explain all our data. According to this model, in the repressed state the same *deoR* repressor complex is bound to both operator sites

joined together through the DNA loop. The increased stability of such a complex is analogous to that of a metal-chelate complex, where the metal binding (*deoR* repressor) is increased because it is held by at least two 'arms' of the chelater (DNA). The model requires the formation of a repressor-protein complex containing two operator binding sites, and a DNA molecule with binding sites which are physically close to each other in the cell.

Distances ranging from 1 to 4 kb have also been used in DNA-ligation studies in which the kinetics of cyclization and of bimolecular reactions were compared¹⁰. The results showed an increase in the effective concentration of sticky ends when they were present on the same molecule. Furthermore, with increasing distance between the ends, this effective concentration decreased in a way very similar to the results presented here.

The formation of protein-DNA loops seems to be a common mechanism for transcriptional regulation in both prokaryotes and eukaryotes^{2,3,5-7,13-15}. The immediate function of the loop seems to be to increase the local concentration of a particular protein in the promoter region resulting in repression (*deo* operon) or activation (enhancers). Regulation from a distance through loop formation is a very adaptable system allowing

targets for the creation of binding sites for different regulatory proteins. This could probably facilitate the evolution of multiple control systems for promoter expression.

This research was supported by grants from the Danish Natural Science Research Council and the Carlsberg Foundation. The technical assistance of Tonny Dedenroth Hansen is gratefully acknowledged.

Received 8 September; accepted 4 December 1986.

1. Irani, M. H., Orosz, L. & Adhya, S. *Gene* **32**, 783–788 (1983).
2. Majumdar, A. & Adhya, S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6100–6104 (1984).
3. Dunn, T. M., Hahn, S., Ogden, S. & Schleif, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5017–5020 (1984).
4. Hahn, S., Dunn, T. & Schleif, R. *J. molec. Biol.* **180**, 61–72 (1984).
5. Dandanell, G. & Hammer, K. *EMBO J.* **4**, 3333–3338 (1985).
6. Valentin-Hansen, P., Albrechtsen, B. & Løve Larsen, J. E. *EMBO J.* **5**, 2015–2021 (1986).
7. Serfling, E., Jasin, M. & Schaffner, W. *Trends Genet.* **1**, 224–230 (1985).
8. Valentin-Hansen, P., Aiba, H. & Schumperli, D. *EMBO J.* **1**, 317–322 (1982).
9. Løve Larsen, J. E., Gerdes, K., Light, J. & Molin, S. *Gene* **28**, 45–54 (1984).
10. Shore, D., Langowski, J. & Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4833–4837 (1981).
11. Mossing, M. C. & Record, M. T. Jr *Science* **233**, 889–892 (1986).
12. Hammer-Jespersen, K. in *Metabolism of Nucleotides, Nucleosides and Nucleobases in Microorganisms* (ed. Munch-Petersen, A.) 203–258 (Academic, London, 1983).
13. Ptasne, M. *Nature* **322**, 697–701 (1986).
14. Griffith, J., Hochschild, A. & Ptashne, M. *Nature* **322**, 750–752 (1986).
15. Takahashi, K. *et al. Nature* **319**, 121–126 (1986).
16. Miller, H. J. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1972).
17. Clark, D. J. & Maaløe, O. *J. molec. Biol.* **23**, 99–112 (1967).
18. Valentin-Hansen, P., Svenningsen, B. A., Munch-Petersen, A. & Hammer-Jespersen, K. *Molec. gen. Genet.* **159**, 191–202 (1978).
19. Munch-Petersen, A. *Eur. J. Biochem.* **6**, 432–442 (1968).
20. Valentin-Hansen, P., Boetius, F., Hammer-Jespersen, K. & Svendsen, I. *Eur. J. Biochem.* **125**, 561–566 (1982).
21. Lowry, O. H., Rosenbrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265–275 (1951).

Structure of co-crystals of tropomyosin and troponin

Steven P. White*, Carolyn Cohen†
& George N. Phillips Jr*‡

* Department of Biochemistry, † Department of Physiology and Biophysics, University of Illinois, Urbana-Champaign, Illinois 61801, USA

‡ Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254, USA

Troponin, a Ca^{2+} -sensitive complex, regulates the motions of tropomyosin on the thin filaments in many muscles. It has three subunits, each with a different architecture and function: TnC binds Ca^{2+} ; TnI binds to actin and inhibits contraction; and TnT binds one complex to each tropomyosin molecule^{1,2}. The troponin complex has an elongated shape with TnC and TnI forming a globular 'head' region and TnT a long (~160 Å) 'tail'³. TnT binds to two widely separated regions of tropomyosin: the head region of the complex is near Cys 190 of tropomyosin and the tail region is near the overlapping joint that links the tropomyosin molecules into filaments⁴. Here we report the X-ray structure determination at 17 Å resolution of glutaraldehyde-treated tropomyosin crystals in which native troponin complex or fragments of TnT have been bound. Our results show that the amino-terminal tail end of TnT spans the head-to-tail joint of the tropomyosin filaments, and that the 'head' region of the whole troponin complex binds ~200 Å away near residues 150–180 of the tropomyosin molecule.

The Bailey crystal form is highly hydrated (95% solvent by volume) and its structure has now been solved to 15 Å resolution including visualization of the head-to-tail joint and correlation of the amino-acid sequence with the electron-density maps^{5,6}. The open meshwork of this lattice (space group $P2_12_12$ with $a = 120$ Å, $b = 244$ Å and $c = 296$ Å) allows large molecules to be incorporated, but also means that the crystals are unusually labile. Thus the troponin complex can sometimes be co-crystal-

lized in the lattice, but leads to large changes in unit cell parameters that have previously prevented determinations of its structure^{7,8}.

We find that treatment of pure tropomyosin crystals with 0.1% glutaraldehyde stabilizes the lattice yet allows the binding of troponin and its components. Diffraction patterns from these glutaraldehyde-treated tropomyosin crystals are virtually identical to those from untreated crystals. The glutaraldehyde reaction is completed before addition of other proteins, which are not crosslinked and bind in a specific manner. Using the structure of pure tropomyosin for initial phases, we have used solvent flattening techniques for phase refinement⁹ including figure-of-merit weighting⁹ of initial and solvent-flattened phases. Subsequent difference Fourier maps were used to locate the troponin in these crystals.

We have layered concentrated solutions of T1, the chymotryptic fragment of TnT of relative molecular mass 19,000 that comprise residues 1–158^{10,11}, in capillaries containing crystals of glutaraldehyde-treated tropomyosin, allowing the fragment to diffuse into the crystal and bind specifically to the filaments. Diffraction data have been collected by screened precession photography. The resulting structure has been solved using difference maps based on the solution of the native tropomyosin structure. The maps reveal the T1 fragment as an elongated mass ~95 Å long and a ~35 Å in width situated on the cross-over regions in the lattice⁶ and extending ~20–30 Å into the 'short arm' (Fig. 1). The fragment thus interacts with a considerable length of the carboxy-terminal end of tropomyosin and extends beyond the head-to-tail joint where it spans a short segment (10–15 amino acids) of the adjacent tropomyosin molecule.

Diffraction data have also been collected from uncross-linked co-crystals grown from a mixture of tropomyosin and the highly α -helical cyanogen bromide-digested CB2 fragment of TnT (residues 71–151)¹². The unit-cell changes arising from the binding of this small fragment are minimal. Difference maps calculated using coefficients generated by subtracting the CB2 co-crystal data from the T1 co-crystal data reveal a distinct peak overlapping the head-to-tail joint of tropomyosin and extending up to the cross-over region of the lattice (Fig. 2). These results indicate that residues 1–70 of troponin T are involved in binding to the termini of tropomyosin and that residues 71–158 extend toward the middle of the tropomyosin molecule.

The entire troponin complex has also been diffused into the cross-linked tropomyosin lattice and visualized. The major features of difference maps calculated from troponin-tropomyosin crystal data minus native tropomyosin data show mainly the mass of the tail region of troponin and no peaks that would correspond to the head portion of the complex. Accordingly, phases of the structure factors for the troponin-tropomyosin crystals were refined using solvent flattening procedures⁶. Difference maps were then calculated using coefficients obtained by subtracting the T1-tropomyosin crystal data to reveal this head region (Fig. 3). The heads appear to be associated with residues 150–180 of tropomyosin, but are bound with lower occupancy than the tail region, and hence are more difficult to visualize.

Although glutaraldehyde was used to stabilize the tropomyosin crystal lattice, our evidence indicates that the TnT binding sites we have identified correspond closely to those of the native molecule. Intensity data from one projection of co-crystals of the CB1 fragment of TnT (residues 1–151) and native tropomyosin (G.N.P. and C.C., unpublished data) were essentially identical to that from the T1-tropomyosin co-crystals. It is likely that only a few tropomyosin lysines react with glutaraldehyde, so that they key linkages holding the TnT to tropomyosin are largely unaffected.

These results define certain of the interactions between troponin and the filaments of tropomyosin. TnT binds along the tropomyosin molecule in an antiparallel fashion, with its amino terminus spanning the carboxy terminus of tropomyosin and a

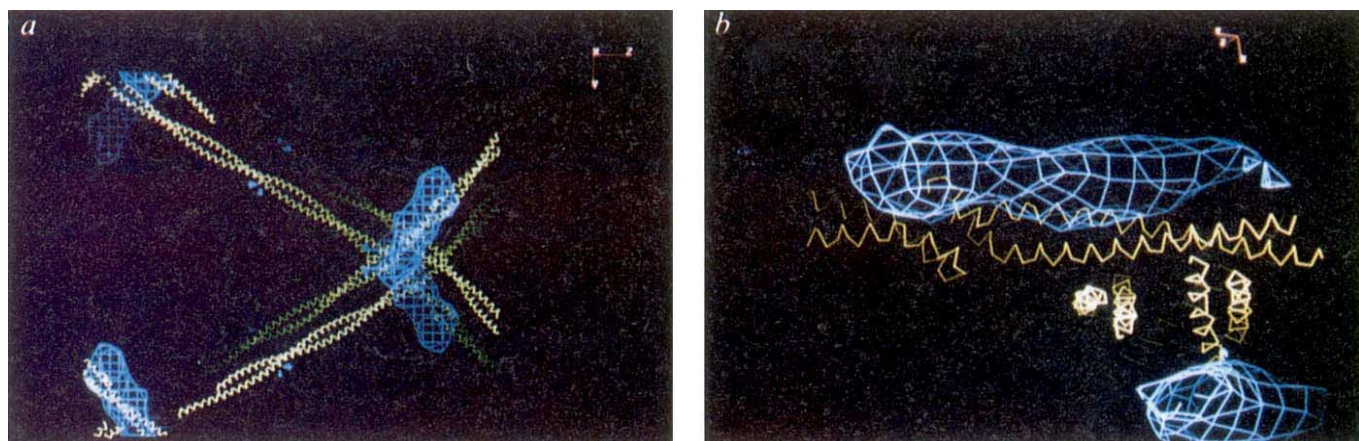


Fig. 1 Fourier map showing differences between tropomyosin-troponin T1 and pure tropomyosin crystals: *a*, view of the whole unit cell showing T1 fragments (blue) binding near the cross-over regions; *b*, different, enlarged view showing binding of T1 extending over the end-to-end joint of tropomyosin (a second symmetry-related T1 fragment is seen at bottom right). Superposed is a model of the α -carbon backbone of the coiled coil (yellow) built to follow the path of the molecule and including known stereochemistry of α helices.

Methods. Tropomyosin was purified as described previously⁵ with the addition of hydroxylapatite chromatography as the final step. The crystals were grown in quartz capillary tubes at or near pH 5.8²³. Troponin was prepared from rabbit skeletal muscle using a modified procedure of Greaser^{1,2} and purified on a Whatman DE52 cellulose anion-exchange column eluted using a linear KCl gradient (adapted from ref. 24). The T1 fragment was prepared directly from troponin by digesting the latter with a 1/750 weight ratio of chymotrypsin (type alpha, Sigma) for 24 min on ice. The digest was fractionated on a Whatman DE52 column. Pure T1 was identified on SDS gel electrophoresis using standards provided by L. Smillie. A 50 mg ml⁻¹ solution of T1 in crystallization buffer was prepared (100 mM sodium acetate, 50 mM ammonium sulphate, pH 5.8) and layered over an α -tropomyosin crystal pretreated with 0.1% glutaraldehyde. Glutaraldehyde pretreatment consisted of preparing a fresh solution of 0.1% glutaraldehyde (Grade 1, Sigma) in crystallization buffer, layering this solution in the capillaries containing the crystals for 2–14 days at 4 °C. Data were collected by screened precession photography to a resolution of 17 Å (745 unique reflections). *R*-factors for merging the data from different film packs were 0.06–0.08. *R*-factors between native and CB2, T1 and Tn data were 0.10, 0.14 and 0.14, respectively. Phases for the difference maps were obtained by combining phases from the native tropomyosin structure⁶ and phases from solvent-flattened maps of each of the troponin-tropomyosin co-crystals. Although the difference maps presented here were calculated using native data from crystals that had not been treated with glutaraldehyde, a control data set from glutaraldehyde-treated native crystals produced virtually identical results. This and subsequent images were calculated and displayed on a Silicon Graphics IRIS 2400 workstation using a modified form of FRODO software (ref. 25 and S. J. Oatley, unpublished data).

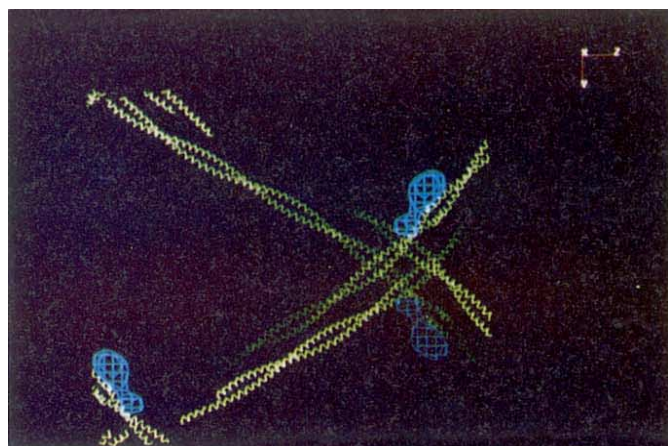


Fig. 2 Difference Fourier map between tropomyosin-troponin T1 and tropomyosin-troponin CB2 crystals. The additional mass present in the T1 crystals seen here reveals that residues 1–70 of TnT bind over the head-to-tail joint of tropomyosin.

small amino-terminal portion of the adjoining molecule. Previous biochemical studies indicated that residues 1–70 of TnT (CB3) interact with TnI and are not involved in binding near the head-to-tail joint of tropomyosin¹³. More recent evidence is consistent with the finding reported here of the interaction between CB3 and the head-to-tail joint¹⁴. TnT also extends ~100 Å toward Cys 190 of tropomyosin but shows little additional mass associated with tropomyosin residues 240–190. The head region of troponin, comprising TnI, TnC and perhaps a small portion of TnT, binds near residues 150–180 of tropomyosin.

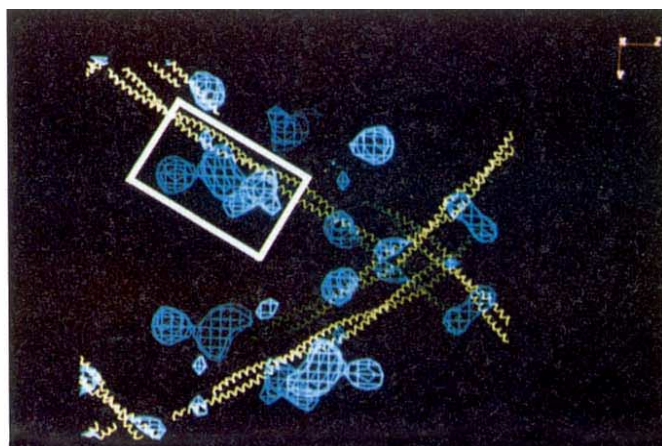


Fig. 3 Fourier map showing differences between tropomyosin-troponin and tropomyosin-troponin T1 crystals. Density corresponding to the head region of troponin is weak, but clearly visible. The white box encloses peaks corresponding to one troponin head. The upper left edge of the box intersects the tropomyosin molecule at residue 120, and the lower right edge at residue 194. Proteins were purified as described in Fig. 1. A 100 mg ml⁻¹ solution of troponin in 10 mM sodium acetate, 1 mM EGTA (ethylenediamine tetraacetic acid), pH 5.8, was prepared and layered over a glutaraldehyde cross-linked tropomyosin crystal, which had been preincubated in the same buffer.

Our results suggest that part of the T1 structure is disordered in the crystal, with only the tropomyosin-binding portions visible in the difference maps. Thus T1 probably comprises about two-thirds of the 185-Å-long TnT subunit³, but only about half the fragment can be seen. Such a picture also accounts for the

lack of connectivity in the difference maps between the head and tail regions of troponin. This two-site model for the attachment of troponin to tropomyosin is in agreement with earlier biochemical studies of binding^{4,15-17} and electron-microscope observations³. The head region of troponin appears to bind weakly to tropomyosin under the ionic conditions used here, and we are studying this interaction further.

Our results reported here together with other findings provide new information about the tropomyosin/troponin switch in muscle. A recent model for regulation suggests that it is primarily the troponin complex that holds tropomyosin extended along the actin helix in the 'off' or resting state⁶. We now suggest that troponin would be bound in a strong invariant link to tropomyosin only near the head-to-tail joint of the filaments. A highly α -helical portion of CB2, although not forming a regular three-chain structure¹⁸, is likely to bind through electrostatic interactions to the two-chain coiled coil of tropomyosin. These interactions are not, however, in regions predicted previously by electrostatic considerations^{19,20}. The middle portion of TnT might form a flexible link between the tightly bound tail region and the head of the troponin complex. The binding of the head region of troponin would depend on the Ca^{2+} saturation of TnC^{16,17,21,22}: at low Ca^{2+} concentration, corresponding to the 'off' state of the switch, the linkages between the subunits are weak, whereas those between TnI and actin, and between TnT and tropomyosin, are relatively strong. In the presence of Ca^{2+} , the head of the troponin complex detaches allowing the movement of tropomyosin to a new position on the thin filament and the binding of force-producing myosin heads. This is the 'on' or 'potentiated' state of the switch (see also ref. 6). Whatever the Ca^{2+} level, the tail region of troponin remains attached to the tropomyosin filament. This invariant linkage strengthens the head-to-tail joint between molecules, enhancing the cooperativity of the thin filament and maintaining the conserved connection of troponin to tropomyosin required for the switching mechanism⁸.

We acknowledge Larry Smillie for initial gifts of troponin T fragments CB1, CB2 and T1, Ed Morris and S. S. Lehrer for suggesting the method of preparation of troponin T fragment T1, J. P. Fillers and D. Stokes for initial diffraction studies and Jack Johnson and Michael Rossmann for use of their film densitometer. This work supported by NIH grants AM32764 (G.N.P.) and AM17346 (C.C.) and a grant from the Muscular Dystrophy Association (C.C.). G.N.P. is an Established Investigator of the American Heart Association.

Received 21 July; accepted 4 December 1986.

- Greaser, M. L. & Gergely, J. *J. biol. Chem.* **246**, 4226-4233 (1971).
- Greaser, M. L. & Gergely, J. *J. biol. Chem.* **248**, 2125-2133 (1973).
- Flicker, P. F., Phillips, G. N. Jr & Cohen, C. *J. molec. Biol.* **162**, 496-501 (1982).
- Mak, A. S. & Smillie, L. B. *J. molec. Biol.* **149**, 541-550 (1981).
- Phillips, G. N. Jr, Lattman, E. E., Cummins, P., Lee, K. Y. & Cohen, C. *Nature* **278**, 413-417 (1979).
- Phillips, G. N. Jr, Fillers, J. P. & Cohen, C. *J. molec. Biol.* **192**, 111-131 (1986).
- Cohen, C., Caspar, D. L. D., Parry, D. A. D. & Lucas, R. M. *Cold Spring Harb. Symp. quant. Biol.* **36**, 205-216 (1971).
- Cohen, C. *et al. Cold Spring Harb. Symp. quant. Biol.* **37**, 287-297 (1972).
- Sim, G. A. *Acta crystallogr.* **13**, 511-512 (1960).
- Ohtsuki, I. *J. Biochem., Tokyo* **86**, 491-497 (1979).
- Tanokura, M., Tawada, Y., Ohoyama, Y., Nakamura, S. & Ohtsuki, I. *J. Biochem. Tokyo* **90**, 263-265 (1981).
- Pearlstone, J. R. & Smillie, L. B. *Can. J. Biochem.* **55**, 1032-1038 (1977).
- Pearlstone, J. R. & Smillie, L. B. *Can. J. Biochem.* **58**, 649-654 (1980).
- Pearlstone, J. R. & Smillie, L. B. *Can. J. Biochem. Cell Biol.* **63**, 212-218 (1985).
- Jackson, P., Amphlett, G. W. & Perry, S. V. *Biochem. J.* **151**, 85-97 (1975).
- Pearlstone, J. R. & Smillie, L. B. *J. biol. Chem.* **257**, 10587-10592 (1982).
- Pearlstone, J. R. & Smillie, L. B. *J. biol. Chem.* **258**, 2534-2542 (1983).
- Parry, D. A. D. *J. molec. Biol.* **146**, 259-263 (1981).
- Nagano, K., Miyamoto, S., Matsumura, M. & Ohtsuki, I. *J. molec. Biol.* **141**, 217-222 (1980).
- Nagano, K. & Ohtsuki, I. *Proc. Jap. Acad. Ser. B* **58**, 73-77 (1982).
- Hitchcock, S. E., Huxley, H. E. & Szent-Gyorgyi, A. T. *J. molec. Biol.* **80**, 825-836 (1973).
- Margossian, S. S. & Cohen, C. *J. molec. Biol.* **81**, 409-413 (1973).
- Caspar, D. L. D., Cohen, C. & Longley, W. J. *J. molec. Biol.* **41**, 87-107 (1969).
- Van Eerd, J. & Kawasaki, Y. *Biochemistry* **12**, 4972-4980 (1973).
- Jones, T. A. in *Computational Crystallography* (ed. Sayre, D.) 303 (Oxford University Press, London, 1982).

Dependence of the mechanical properties of actin/ α -actinin gels on deformation rate

Masahiko Sato, William H. Schwarz* & Thomas D. Pollard

Departments of Cell Biology and Anatomy and of * Chemical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

The cortical cytoplasm, including the cleavage furrow, is largely composed of a network of actin filaments that is rigid even as it is extensively deformed during cytokinesis^{1,2}. Here we address the question of how actin-filament networks such as those in the cortex can be simultaneously rigid (solid-like) and fluid-like. Conventional explanations are that actin filaments rearrange by some combination of depolymerization and repolymerization; fragmentation and annealing of filaments; and inactivation and re-establishment of crosslinks between filaments³⁻⁵. We describe the mechanical properties of a model system consisting of actin filaments and *Acanthamoeba* α -actinin⁶⁻⁸, one of several actin crosslinking proteins found in amoeba and other cells^{4,9}. The results suggest another molecular mechanism that may account for the paradoxical mechanical properties of the cortex. When deformed rapidly, these mixtures are 40 times more rigid than actin filaments without α -actinin, but when deformed slowly these mixtures were indistinguishable from filaments alone. These time-dependent mechanical properties can be explained by multiple, rapidly rearranging α -actinin crosslinks between the actin filaments, a mechanism proposed by Frey-Wyssling¹⁰ to account for the behaviour of cytoplasm long before the discovery of cytoplasmic actin or α -actinin. If other actin-filament crosslinking proteins behave like *Acanthamoeba* α -actinin, this mechanism may explain how the cortex recoils elastically from small rapid insults but deforms extensively when minute forces are applied over long periods of time^{1,11,12}.

We evaluated the mechanical properties of mixtures of actin and *Acanthamoeba* α -actinin in a cone-and-plate rheometer using a sinusoidal pattern of low amplitude deformations^{8,13-15}. As described previously¹⁵, the small strains (deformations) used do not alter the mechanical properties of equilibrated samples of actin.

It is known that α -actinin increases both the rigidity (elasticity, G') and viscosity (η') of actin filament⁸, but we show here that the extent of these changes depends strongly on the rate of deformation (Fig. 1). At equilibrium, actin filaments alone at 24 μM form a weak gel (viscoelastic solid) characterized by a frequency-independent rigidity (equilibrium elastic modulus), $G_e = 14 \pm 4 \text{ dyn cm}^{-2}$ (Fig. 1a), as shown previously¹⁵. With the addition of α -actinin, both the dynamic elasticity G' and dynamic viscosity η' of rabbit actin filaments are elevated 40-fold when sinusoidally deformed (strained) at 1.0 Hz. But at low frequencies ($<10^{-3} \text{ Hz}$), both parameters approach values for actin filaments alone (Fig. 1a).

The material properties of mixtures of polymerized actin and α -actinin at high rates of sinusoidal deformation are attributable to multiple crosslinks between the filaments, as control experiments exclude substantial contributions by the other components in the mixture. First, free α -actinin, free actin filaments and free nonfilamentous actin do not contribute significantly to the mechanical properties of the gel because G' and η' for each alone are at least 10-fold lower at high frequencies ($>10^{-2} \text{ Hz}$) (Fig. 1). Second, complexes of α -actinin with actin monomers (if any) do not contribute, because α -actinin does not alter the rheological properties of nonfilamentous actin alone (Fig. 1b). Third, the temperature dependence of the elasticity of the mixture of α -actinin with actin filaments is opposite to that of the

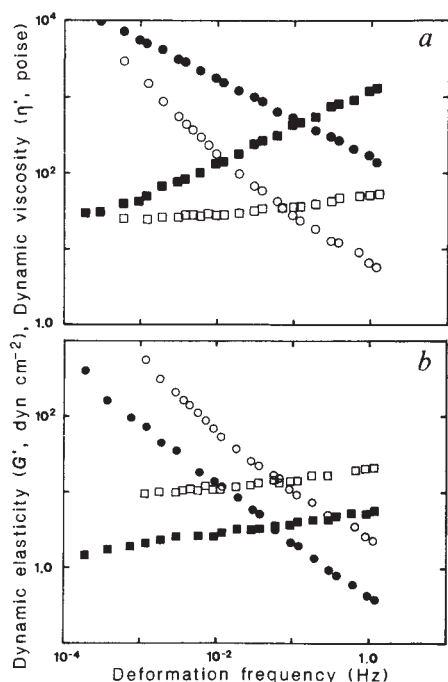


Fig. 1 Effect of *Acanthamoeba* α -actinin on the mechanical properties of actin over a wide range of deformation rates (frequencies). *a*, 24 μ M filamentous actin alone (F-actin; white squares, G' ; white circles, η') or filamentous actin plus 1.6 μ M α -actinin (F-actin + α -actinin, black squares, G' ; black circles, η'). *b*, 1.6 μ M α -actinin in polymerization buffer (black squares, G' ; black circles, η') and 24 μ M nonfilamentous actin plus 1.6 μ M α -actinin (white squares, G' ; white circles, η') in low-salt buffer (0.1 mM MgCl_2 , 1 mM EGTA, 2 mM PIPES pH 7.0, 0.2 mM ATP, 1 mM dithiothreitol (DTT), 0.5 mM NaN_3).

Methods. *a*, G' and η' were evaluated with a cone-and-plate rheometer (Weissenberg rheogoniometer) for the frequency range 10^{-4} –3 Hz. The sample was deformed by sinusoidally oscillating the bottom plate relative to a freely suspended cone using small displacements that do not disrupt the equilibrium mechanical properties of actin filaments¹⁵. From the phase and amplitude response of the cone relative to the plate, we calculated G' and η' using linear viscoelastic theory^{15,19}. In this type of experiment, the freely suspended cone oscillates in phase with the plate if a solid is present between cone and plate. If a newtonian liquid like water is placed between cone and plate, the cone oscillates -90° out-of-phase with the plate. For a complex material like actin that exhibits both solid-like and liquid-like properties, dynamic elasticity and dynamic viscosity are the in-phase and -90° out-of-phase components of the cone response. These dynamic parameters measure elasticity and viscosity under nondisruptive conditions and have no relation to shear rate dependent parameters such as shear viscosity^{15,19}. Gel-filtered actin from rabbit skeletal muscle³⁰ and *Acanthamoeba* α -actinin^{6,7} were equilibrated in polymerization buffer (100 mM KCl, 1 mM MgCl_2 , 1 mM EGTA, 10 mM PIPES pH 7.0, 0.2 mM ATP, 1 mM DTT, 0.5 mM NaN_3) for 1,000 min at 25 $^\circ\text{C}$ in the rheometer¹⁵. The rheometer cone and plate were 10 cm in diameter with a cone angle of 0.01729 rad. The deformation frequency was varied with a constant amplitude of 0.0046 rad for a maximum strain of 0.083¹⁵. Below a frequency of 6×10^{-2} Hz, actin filaments have a frequency-independent elasticity G_e (see text), characteristic of a viscoelastic solid^{15,19}. At 1.0 Hz, α -actinin increases both the G' and η' by 40-fold, but at 6×10^{-4} Hz the mixture was mechanically indistinguishable from actin filaments alone. These data were reproduced for two sets of experiments with two preparations of α -actinin. Both types of samples contained numerous actin filaments shown by electron microscopy³¹. *b*, G' tends toward zero at low frequency indicating that α -actinin alone is a viscoelastic liquid. G' and η' are identical for α -actinin in polymerization buffer and low-salt buffer. For the mixture of nonfilamentous actin and α -actinin, G' and η' are the same as values for nonfilamentous actin alone indicating that α -actinin does not significantly affect the mechanical properties of nonfilamentous actin for this frequency range.

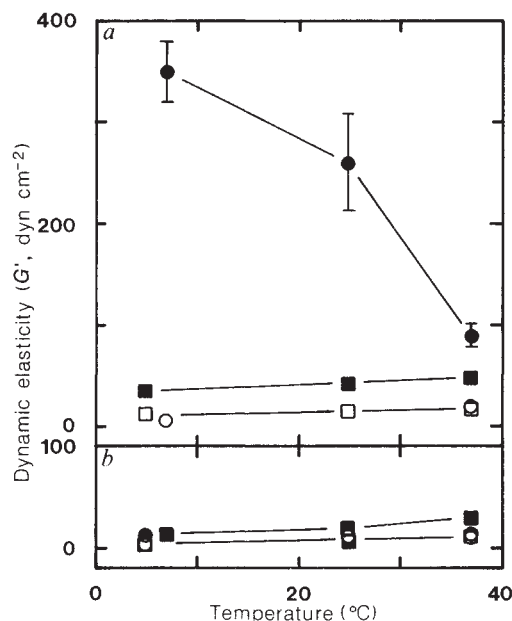


Fig. 2 Temperature effects on dynamic elasticity. 24 μ M filamentous actin (G' , black squares), 24 μ M filamentous actin plus 1.6 μ M α -actinin (black circles), 24 μ M nonfilamentous actin (white squares), 24 μ M nonfilamentous actin plus 1.6 μ M α -actinin (white circles). These symbols correspond to the calculated means for 2–5 sets of experiments with proteins from different samples of *Acanthamoeba*. Standard deviations were smaller than the space occupied by the symbols, except where designated by error bars. *a*, Deformation frequency = 6×10^{-1} Hz. The mixture of filamentous actin plus α -actinin had higher values of G' and an opposite trend of G' as a function of temperature compared with the other samples, showing that functional crosslinks between actin filaments dominate the mechanical properties for the mixture at high frequencies. *b*, Deformation frequency = 6×10^{-4} Hz. For all samples, G' increased with temperature in a similar fashion.

other components (actin monomers + α -actinin and actin filaments alone) in the sample (Fig. 2).

The key experiment to elucidate the frequency dependence of the mechanical properties of these gels was to measure the equilibrium binding constant for α -actinin to actin filaments (Fig. 3), so that we could estimate the steady-state association and dissociation rates of the crosslinks. In sedimentation binding experiments, the dissociation constant K_D of the actin filament and α -actinin complex is 26 μ M (Fig. 3). From this value we estimated the number of α -actinin molecules bound and the possible exchange rates. First, under the conditions used in Figs 1 and 2, only 35% of the total α -actinin (1.6 μ M) is bound to actin filaments; but as the average filament length is probably $>5 \mu$ m (ref. 16) >50 α -actinins are bound to each filament at any instant. Second, if the association rate constant is assumed conservatively to be $10^5 \text{ M}^{-1} \text{ s}^{-1}$, the dissociation rate constant is $>2 \text{ s}^{-1}$. The actual value may be one or two orders of magnitude higher, because other proteins that bind to actin filaments such as myosin have association rate constants of 10^6 to $10^7 \text{ M}^{-1} \text{ s}^{-1}$. Consequently, at equilibrium the α -actinin in these gels binds to and dissociates from the actin filaments rapidly (2 – 200 s^{-1}), like the weakly bound intermediates in the actomyosin crossbridge cycle in muscle contraction^{17,18}. The difference with myosin, however, is that α -actinin 'cycling' is independent of ATP.

The rapid binding equilibrium of α -actinin with actin filaments means that crosslinks between filaments can rearrange quickly and together with the existence of multiple crosslinks between filaments explains why the mechanical properties of the gel depend on the rate of sinusoidal deformation (Fig. 1).

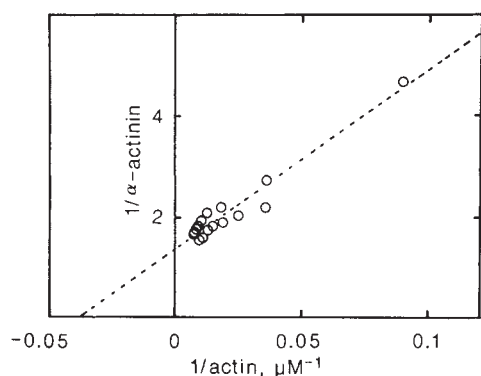


Fig. 3 Dissociation constant for the complex of *Acanthamoeba* α -actinin with actin filaments.

Methods. We incubated 8.5 nmol of α -actinin with 408 nmol [14 C]iodoacetamide (NEN, Boston) for 90 min in 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1 mM DTT in 1 ml total volume at room temperature. The reaction was stopped with 3 μ mol DTT and the products were concentrated to 0.5 ml vol. by dialysing against Aquacide II (Calbiochem, La Jolla) for 60 min, 5 °C. The products were gel filtered on a S-300 column (1.2 $\times$ 47 cm) pre-equilibrated with Tris-EDTA-DTT buffer and radioactivity counted on a Beckman LS 7000 liquid scintillation counter. The labelling varied from 20 to 100% for three batches of α -actinin. The gelation activity was assayed with the 'falling ball' technique³² and varied from 80 to 150% compared with unlabelled α -actinin. Iodoacetamide-labelled α -actinin (1.6 μ M) was incubated for 1,000 min at 25 °C with various concentrations of rabbit muscle actin (5–180 μ M) in polymerization buffer (100 mM KCl, 1 mM MgCl₂, 1 mM EGTA, 10 mM PIPES pH 7.0, 0.2 mM ATP, 1 mM DTT, 0.5 mM NaN₃). The mixtures were centrifuged for 25 min at $1.6 \times 10^5 g$ in an Airfuge (Beckman) to pellet the actin filaments and the supernatants (70 μ l) were assayed for radioactivity to measure the free α -actinin. The amount of α -actinin in the pellet was calculated from the original total and the data are presented as a double reciprocal plot by the method of Brenner *et al.*¹⁷. These data from two different batches of α -actinin were fit to a linear regression line. The intercept on the x axis gives $K_D = 26 \mu$ M.

When force is applied slowly, the α -actinin crosslinks rearrange faster than the displacement of actin filaments relative to each other, so they offer little or no resistance. When force is applied more rapidly than the whole population of crosslinks can rearrange, there will be physical connections between the filaments that make the network rigid. Amazingly, Frey-Wyssling¹⁰ in 1948 proposed a similar model without knowledge of cytoplasmic actin or crosslinking proteins.

Because association and dissociation rates are functions of temperature, the rate of crosslink re-arrangement should also increase with the temperature. This may explain why the rigidity (G') of the gel mixture decreases with temperature, first shown by Abe and Maruyama⁸, whereas G' for the individual components increases with temperature (Fig. 2a). During low-frequency deformations this effect of temperature on the mixture is not observed because the rate of crosslink re-arrangement is faster than the deformation frequency (Fig. 2b). We have recently confirmed every aspect of this model for α -actinin in fluorescence photobleaching recovery experiments (M.S., D. Loftus, J. Cooper, E. Elson, C. Freider and T.P., in preparation).

The effects of α -actinin on the mechanical properties of actin filaments were unexpected because, to our knowledge, similar effects have not been observed for either synthetic polymers or isolated proteins. Addition of covalent crosslinkers to a viscoelastic solid of synthetic polymers results in the elevation of the equilibrium elastic modulus (G_e) rather than only an increase in G' at high frequency¹⁹. Although actin-filament crosslinking

proteins such as α -actinin, actin-binding protein, spectrin and others^{4,9} do not covalently link filaments, previous experimental and theoretical studies of filament networks have assumed for simplicity that the crosslinks are permanent^{20–23}. In dynamic models of cytoskeletal gels, for example, crosslinking proteins have been modelled as permanent links that are disrupted by regulatory factors such as calcium^{22,23}. Our data show that the network of *Acanthamoeba* α -actinin and actin filaments is both rigid and deformable depending on the rate of sinusoidal deformation without requiring regulation by inactivating crosslinks, fragmenting filaments or depolymerizing filaments. This may be particularly important for *Acanthamoeba* α -actinin, as it is not regulated by calcium⁷.

Macrophage α -actinin²⁴ and actin-binding protein²⁵ are also likely to be in rapid equilibrium with actin filaments because they bind relatively weakly ($K_D = 0.2–0.5 \mu$ M), especially α -actinin in the presence of calcium ($K_D = 1 \mu$ M). As discussed above, the dissociation rate constants may be in the range of $0.02–10 s^{-1}$. Consequently, the mechanical properties of actin filaments with these crosslinkers should also depend on the rate of deformation. This notion of dynamic crosslinks is consistent with the recent data of Zaner^{14,26} and should become more obvious at lower and lower frequencies.

The dynamic model¹⁰ for α -actinin crosslinking of actin filaments may be applicable to the amoeba cortex and may explain some features of cytokinesis. Like amoebas^{7,27} both sea urchin eggs²⁸ and cultured vertebrate cells²⁹ have a cortical cytoplasm that is rich in actin filaments and α -actinin. Hiramoto¹ and Peterson *et al.*¹² showed that the cortex of sea urchin eggs and fibroblasts is rigid (elastic) and more resistant to rapid deformations than to slow deformations. For example, although the cleavage furrow generates only minute forces of 3×10^{-3} dyn or stresses between 10 and 50 dyn cm⁻² over 10–30 min (between 10^{-3} and 10^{-4} Hz), it is capable of grossly deforming the highly rigid cortex¹. Our data suggest that α -actinin localized to the cortex^{27–29} forms an actin gel that resists rapid insults yet deforms with minimal resistance provided that stress is applied for a long time.

Supported by NIH grants GM 26338 and GM 26132 (T.D.P.) and a Postdoctoral Fellowship from the Muscular Dystrophy Association (M.S.).

Received 14 October; accepted 10 December 1986.

- Hiramoto, Y. *Biorheology* **6**, 201–234 (1970).
- Bray, D., Heath, J. & Moss, D. *J. Cell Sci. Suppl.* **22**, 1–18 (1986).
- Taylor, D. L. & Condeelis, J. S. *Int. Rev. Cytol.* **56**, 97–144 (1979).
- Stossel, T. P. *et al. Rev. Cell Biol.* **1**, 353–402 (1985).
- Taylor, D. L. & Fecheimer, M. *Phil. Trans. R. Soc. B* **288**, 185–197 (1983).
- Pollard, T. D. *J. biol. Chem.* **256**, 7666–7670 (1981).
- Pollard, T. D. *et al. Cell Motil.* (in press).
- Abe, S. & Maruyama, K. *J. Biochem., Tokyo* **73**, 1205–1210 (1973).
- Pollard, T. D. & Cooper, J. A. *Rev. Biochem.* **55**, 987–1035 (1986).
- Frey-Wyssling, A. *Submicroscopic Morphology of Protoplasm and its Derivatives*. (Elsevier, New York, 1948).
- Sato, M., Wong, T. Z., Brown, D. & Allen, R. D. *Cell Motil.* **4**, 7–23 (1984).
- Petersen, N. O., McConaughy, W. B. & Elson, E. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5327–5331 (1982).
- Zaner, K. S. & Stossel, T. P. *J. Cell Biol.* **93**, 987–991 (1982).
- Zaner, K. S. & Stossel, T. P. *J. biol. Chem.* **258**, 11004–11009 (1983).
- Sato, M., Leimbach, G., Schwarz, W. H. & Pollard, T. D. *J. biol. Chem.* **260**, 8585–8592 (1985).
- Lanni, F. & Ware, B. R. *Biophys. J.* **46**, 97–110 (1984).
- Brenner, B., Schoenberg, M., Chalovich, J. M., Greene, L. E. & Eisenberg, E. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7288–7291 (1982).
- Taylor, E. W. *CRC Crit. Rev. Biochem.* **6**, 103–164 (1979).
- Ferry, T. D. *Viscoelastic Properties of Polymers* 11–118 (Wiley, New York, 1970).
- Stokke, B., Mikkelsen, A. & Elgsaeter, A. *Biophys. J.* **49**, 319–326 (1986).
- Schanus, E., Booth, S., Hallaway, B. & Rosenberg, A. *J. biol. Chem.* **260**, 3724–3730 (1985).
- Oster, G. F. & Odell, G. M. *Cell Motil.* **4**, 469–503 (1984).
- Nossal, R. *Polymer Preprints* **27**, 241–242 (1986).
- Bennett, J. P., Zaner, K. S. & Stossel, T. P. *Biochemistry* **23**, 5081–5086 (1984).
- Hartwig, J. H. & Stossel, T. P. *J. molec. Biol.* **145**, 563–581 (1981).
- Zaner, K. S. *J. biol. Chem.* **261** (1986).
- Brier, J., Fecheimer, M., Swanson, J. & Taylor, D. L. *J. Cell Biol.* **97**, 178–185 (1983).
- Mabuchi, I. *et al. J. Cell Biol.* **100**, 375–383 (1985).
- Fujiwara, K., Porter, M. E. & Pollard, T. D. *J. Cell Biol.* **79**, 268–275 (1978).
- MacLean-Fletcher, S. & Pollard, T. D. *Biochem. biophys. Res. Commun.* **96**, 18–27 (1980).
- Jockusch, B. M. & Isenberg, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3005–3009 (1981).
- MacLean-Fletcher, S. & Pollard, T. D. *J. Cell Biol.* **85**, 414–428 (1980).

Mapping using gene encyclopaedias

D.L. Daniels and F.R. Blattner

Overlap cloning and large fragment electrophoresis bring total knowledge of the *E. coli* genome into prospect.

THE circular genome of *Escherichia coli* consists of 4.7 million basepairs $\pm$ 100 kilobase pairs, corresponding to a genetic map of 100 minutes. Over 1,000 genes are mapped¹, and perhaps 3,000 remain to be discovered. Recently, a restriction map has been determined using *NotI* and *SfiI*, having 23 and 27 cleavage sites, respectively (C.L. Smith and C.R. Cantor, Columbia University, personal communication.) This map provides the most accurate measurement of the genome size.

During the past five years our laboratory has been working on a programme to create an ordered set of DNA clones spanning the genome of *E. coli*. The term "gene encyclopaedia" is suggested for this organized clone bank to distinguish it from a "library", which is a random collection. Similar efforts have been initiated by Olson *et al.*² for yeast, and by Coulson *et al.*³ for the nematode. George Brownlee at Oxford has also recently started an *E. coli* cosmid project along similar lines (personal communication).

Utility

The *E. coli* gene encyclopaedia, when completed, will provide researchers with important advantages. For example, if the map position is known for a desired gene it could be "cloned by phone". If not, it could still be isolated more quickly by screening an encyclopaedia than a library, as fewer individual clones would need to be screened to ensure coverage of the entire genome. Moreover, a mutation with a phenotype that cannot be detected in clones could be genetically mapped using the encyclopaedia clones in marker rescue tests, thereby identifying a clone of the gene.

The encyclopaedia could also be used to study gene regulation in *E. coli*. For example, to identify all the genes that are under the control of a particular regulatory mechanism such as catabolite repression, the SOS pathway or heat shock, the encyclopaedia could be screened with RNA probes synthesized in *E. coli* in response to the inducing stimulus. Once identified, the clones could be used for quantitative studies. Regions lacking regulated genes could also be demonstrated. In the past, characterization of pleiotropic regulatory systems has required individual testing of postulated targets.

Finally, the encyclopaedia will aid in physical mapping of the genome, including its sequencing, and facilitate engineering modification. Approximately eight

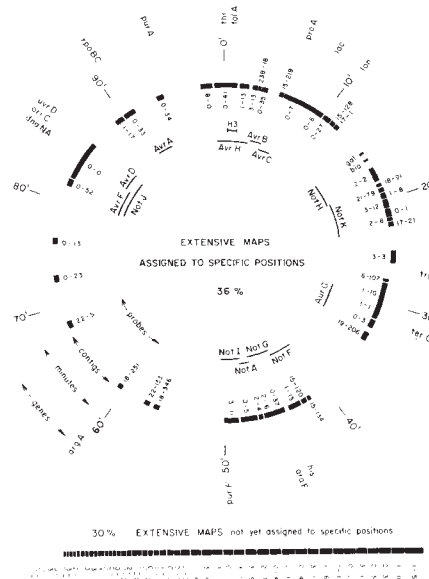


Fig. 1 Sizes and positions of located contigs are indicated on the circular map. Extensive maps which have not yet been located are shown below the circle.

per cent of *E. coli* has been sequenced so far through uncoordinated individual efforts.

Overlap cloning

E. coli K-12 strain MG1655 was chosen as the reference strain after consultation with Barbara Bachman of the Yale University Microbial Stock Center, because it is prototrophic, has not been artificially mutagenized, and lacks gamma delta sequences, which are known to promote DNA rearrangement.⁴

A minimal encyclopaedia would require about 125 cosmid clones or 250 phage lambda clones. Lambda vectors were chosen despite their smaller capacity, out of concern for accuracy and completeness. In general, cosmid and plasmid clones may accumulate deletions as their rate of doubling depends inversely on DNA length, and because expression of cloned genes may be selected against due to deleterious metabolic imbalances. This could be especially true in the more useful high copy number cosmids.

Phage lambda, on the other hand, destroys its host in ten minutes, and thus need not coexist with it. Moreover, there is inherent selection against deletion in lambda clones due to the existence of a lower limit on the size of DNA that can be propagated. Charon 28⁵ and Charon 40⁶ were used. The latter vector has a 24 kilobase pair DNA capacity and offers a very con-

venient polystyrene method for isolation of vector arms.⁶

The initial approach resembled the 'random' method used for DNA sequencing. Three types of libraries were made from partial digests of *E. coli* DNA with *Bam*HI, *Eco*RI, and *Hind*III. CsCl gradient centrifugation of the packaged phage enriched the libraries for large DNA inserts. Individual clones were chosen at random, cut with each of the three enzymes and pairs of enzymes, and mapping gels were run. Fragment sizes were measured using a digitizer and marker calibration program. The results were fed into a computer which was programmed to search for overlaps between clones.

The 'signature' used to identify overlaps consisted of a list of fragment sizes for each clone plus the binary data of whether the fragment was cut by each of the other enzymes. The program scanned the database for pairs of clones sharing at least two signature elements and overlapping maps were assembled from amongst the possibilities by hand.

Selection of the signature was tricky. A prime consideration was the desire to obtain a complete *Bam*HI-*Eco*RI-*Hind*III map of *E. coli* as a byproduct. The order and position of restriction sites in most of the clones was determined, but some of the maps were indeterminate. Thus the signature used by the program was an unordered one that could be obtained from the data regardless of the state of the mapping. The disadvantage of this was that signature elements in the middle of the cloned insert were not distinguishable from those at the ends, so the strongest matches corresponded to the less interesting duplicate clonings, whereas the interesting partial overlaps had to be sorted out from the noise of chance fragment identities.

Although a 'denser' unordered signature such as that of Coulson *et al.*³ might be beneficial, in retrospect an ordered signature would have been much more powerful. A set of partial restriction enzyme digests run by the Smith and Birnstiel⁷ method in a field inversion gel could be a good choice for future efforts.

Data for 2,000 clones has been coalesced into 68 overlapping contiguous regions or 'contigs' ranging in size from 18 to 200 kilobase pairs, accounting for 67 per cent of the genome. Much of the remainder is probably covered by the over 250 unique unattached clones that remain in the collection.

Field inversion gels

The advent of orthogonal field⁸ and field inversion⁹ gel electrophoresis for separating large DNA molecules allowed a quantum leap in assembly of the encyclopaedia. Using field inversion gels DNA fragments of 100 to 500 kilobase pairs were isolated from restriction digests with *NotI*, *SfiI*, *SpeI*, *AvrII* and *HindIII*. These were used as hybridization probes to spot blots of the entire clone bank. The resulting data helped order the collection by identifying neighboring contigs. In many cases this allowed the discovery of new overlaps. For example, overlaps of single fragments could often be identified in the context of a few hundred kilobase pair subregions. This approach also allowed the sizes of remaining gaps to be estimated by subtraction. The positions of the gaps could often be pinpointed as well. In general, gaps were small, and we suspect in many cases that contigs abut without an overlapping clone to prove it.

Genetic map assignment

The literature was surveyed for restriction maps of known genes, and a computer program located the corresponding clones. Other clones were located by genetic tests such as the elucidation of those carrying *lacZ* by their formation of blue plaques, and the ability of lambda clones carrying the wild type allele to plate on *groE*, *dnaK* and *dnaJ* bacteria. Tests for other markers prove to be laborious, however, and have not been used extensively.

Figure 1 shows the state of the encyclopaedia as of 30 January 1987. The genetically assigned clones are positioned around the circle, and correspond to 36 per cent of its length. The ones across the bottom depict unassigned extensive contigs, corresponding to 30 per cent of the genome. It is to be expected that their assignment by hybridization will be accomplished soon.

After all the isolated clones have been assigned, gaps will remain, possibly due to awkward restriction site placement. It will be fascinating to examine the gaps that remain after the trivial gaps have been filled, for these nonclonable regions may carry biologically interesting structures.

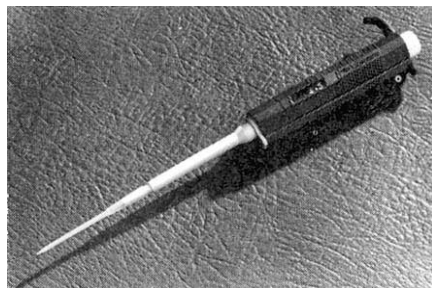
Donna L. Daniels and Frederick R. Blattner are at the Department of Genetics, University of Wisconsin, Madison, Wisconsin 53705, USA.

1. Bachmann, J. J. *microbiol. Rev.* **47**, 180-230 (1983).
2. Olson, M. V. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7826-7830 (1986).
3. Coulson, A., Sulston, J., Brenner, S. & Karn, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7821-7825 (1986).
4. Guyer, et al. *Cold Spring Harb. Symp. quant. Biol.* **45**, 135 (1981).
5. Rimm, D., Horness, D., Kucera, J. & Blattner, F. R. *Gene* **12**, 301-309 (1980).
6. Dunn, I. & Blattner, F. R. *Nucleic Acids Res.* (in the press).
7. Smith, H. O. & Birnstiel, M. L. *Nucleic Acids Res.* **3**, 2387-2398 (1976).
8. Schwartz, D. & Cantor, C. R. *Cell* **37**, 67-75 (1984).
9. Carle, G. F., Frank, M. & Olson, M. V. *Science* **232**, 65-68 (1986).

ASM anecdotes

This week some of the products that will be on exhibit at the American Society for Microbiology meeting in Atlanta, Georgia, 1-6 March, are previewed.

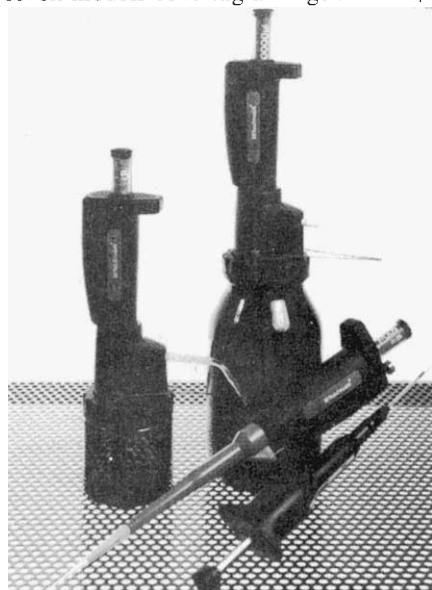
FROM LabSystems Ltd comes a small volume pipette suitable for handling small amounts of enzymes, in applications such as the addition of restriction enzymes in biotechnological procedures, and the loading of electrophoresis gels (Reader Service No. 101). The £90 (UK) Digital



Pipettes extremely small volumes.

Finnpipette is adjustable for volumes between 0.5 µl and 10 µl, and the digital readout makes delivery of the exact amount of substance easy. No special tips are required for the Digital Finn timer: it uses LabSystems' Universal tips, that can be used with other pipettes with volumes of up to 200 µl. Computerized calibration documents are provided with every model. LabSystems' products can be seen in booth numbers 1810 and 1811.

Whatman LabSales Ltd is now selling liquid handling products, with everything from air and positive displacement pipettors, to dispensers and peristaltic pumps (Reader Service No. 102). The range of air displacement pipettors consists of seven models covering a range from 1 µl



A view of Whatman's products for liquid handling and dispensing.

to 5,000 µl, and each has precise volume adjustments, a body designed to prevent variations due to heat transfer during prolonged use, and a range of accessories including pipette tips and a stand. The colour-coded digital positive displacement pipettors cover a range of volumes from 2 µl to 1,000 µl, and are useful for viscous solutions, or in cases where sterilization by autoclaving is necessary. For larger volumes of 50-10,000 µl, four digital dispensers, each supplied with two amber reservoirs, are available. Information on Whatman LabSales' new line can be found in booth numbers 1647 and 1648.

The Nalgene Filling Bell from Nalge, in booths 1049 and 1050, is designed for the transfer of sterile media aseptically from large containers to media bottles or culture vessels (Reader Service No. 103). The \$12 (US) filling bell is transparent and



Nalge's filling bell fits most bottles.

autoclavable, and is made of durable, nontoxic polycarbonate, offering a safe alternative to glass. The bell has an internal diameter of 70 mm, fits all common media bottles, and its autoclavable polypropylene tubing adapter accepts 0.25-0.50 inch tubing.

Growing and detecting

University Micro Reference Laboratory, Inc. sells a wide range of media suited for the growth of a myriad of different microorganisms (Reader Service No. 104). Now its 150 different media for bacteriological, biotechnological and mycological applications are available in preweighed PacketMedia pouches, obviating the need to buy and keep large amounts of specific

media. PacketMedia pouches come in sizes for the production of 500 ml, 1 litre and 2 litres of prepared media, and most have a half-life of four to five years. Information on the range of University Micro Reference Laboratory's media can be found in its booth, number 1752.

Latex agglutination tests are available from Wellcome Diagnostics, who will be exhibiting in island 7D (Reader Service No. 105). The tests work by using latex



Detects bacteria by visible agglutination.

particles coated with group-specific antibody, and show results (visible agglutination) when the corresponding bacterial antigen reacts with the particles. Wellcome's tests, ranging in price from \$77–205 (US), allow detection of A, B, C, D, F, and G *Streptococci*, *Staphylococcus aureus*, *H. influenzae*, *S. pneumoniae*, and *N. meningitidis*.

Flow Laboratories has introduced a new range of cell growth products manufactured by Collaborative Research, Inc. (Reader Service No. 106). Included in the range are broad purpose mitogens, biomatrix products, formulated medium specialties, and serum-free media additives. These products are provided in sterile, lyophilized pack sizes for propagation and differentiation of human and animal cells in culture. Flow Laboratories' cell culture products will be in island 7C and booth numbers 849 and 850.

Fermenting focus

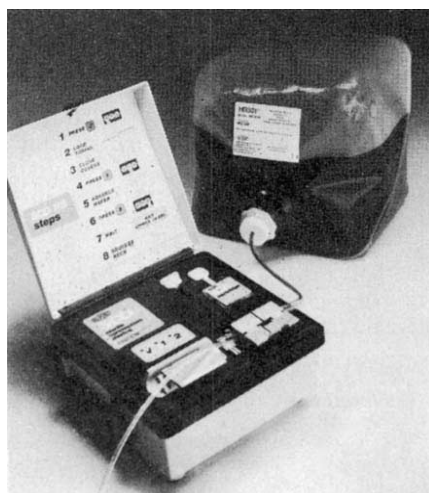
Charles River Biotechnological Services will have information on its 5200R Opticell Culture System in booth numbers 1505 and 1506. (Reader Service No. 107). The \$28,750 (US) system is a benchtop version of the larger Opticell 5300 and 5400, and is designed to produce monoclonal antibodies and other biologicals in research quantities, to serve as a process development tool. The 5200R exhibits real-time feedback, and is suitable for the culture of attachment-dependent or suspension cells. The Opticell 5200R has several features not found on the larger systems. The system is available with a dual loop option, allowing two complete bioprocess shelves to share the same pro-



A controlled culture system that fits right on the benchtop.

cess controller, for a truly controlled experiment without the added cost of having two separate culture systems. The operator also has independent control of the parameter settings for each loop, such as dissolved O_2 , pH, temperature and feed/harvest rates. The system is self-contained, with a clean air curtain to avoid contamination.

Designed to make rapid sterile connections in cell culture or fermentation systems, the Du Pont SCD-IIB sterile connection device is also compact and portable (Reader Service No. 108). Using the \$9,450 (US) SCD-IIB, it is possible to add liquids or media, sample fermenter and receiver contents, replace fermenter com-



Du Pont's SCD-IIB really makes the connection.

ponents, harvest products, or transfer cells between containers, in a few steps. Du Pont's sterile connection device operates by providing sterile butt welds between ends of thermoplastic tubing, protecting the integrity of bioreactor systems, and Du Pont claims its device has never failed in a test of 3,000 connections of 300 individual cultures. Du Pont will be exhibiting its wares in island 13C.

Fisher Scientific's new line of laboratory circulators all share a common feature, according to Fisher: a high performance to price ratio (Reader Service No. 109). The line includes six immersion, three heating, and six refrigerated units,

ADVERTISEMENTS

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5–10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.47

WESTERN BLOTS IN MINUTES

Our new GENIE electrophoretic blotter incorporates space-age plate electrodes for a uniform, high-intensity field. It is easy to load, and uses less than a liter of buffer, yet can blot gels up to 15 by 17 cm. in as little as ten minutes.

only \$190



Idea Scientific Company
P.O. Box 2078
Corvallis, Oregon 97339
(503) 758-0999

ASM Booth 1715

Reader Service No.14

MICRO-DISSECTION?

MICRO-INJECTION?

MICRO-BIOLOGY?

MICRO-TECHNOLOGY?

MICRO-MANIPULATOR?

MK I

THE ORIGINATOR

SINGER

DETAILS FROM:

SINGER INSTRUMENT CO LTD
TREBOROUGH LODGE,
ROADWATER, WATCHET,
SOMERSET TA23 0QL ENGLAND.
TEL: (0984) 40226
TELEX 337492 COMCAB G

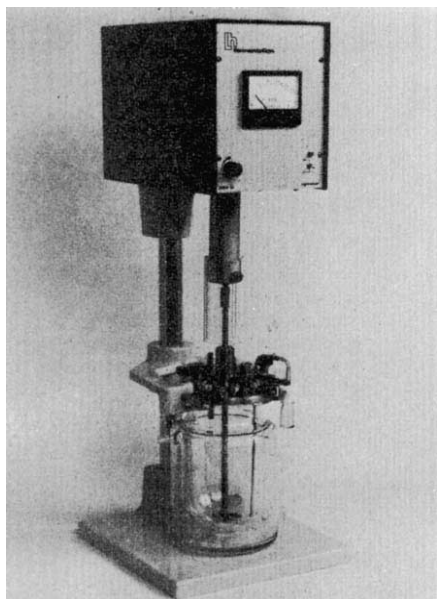
Reader Service No.50

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

available for \$350–\$1875 (US), \$760–\$1135 (US), and \$1325–\$2395 (US), respectively. The units offer temperature control to $\pm 0.02^\circ\text{C}$ via proportional heating, user-adjustable pre-set temperatures, stainless-steel components, resettable circuit breakers, 1,000 W heaters, and selectable pump speeds of between 7 and 15 litres per minute. Fisher Scientific will have its range of products for the microbiology laboratory in booth numbers 1010, 1011 and 1052–1055.

For the cultivation of shear-sensitive cells, LH Fermentation, in booth numbers 1856 and 1857, sells a direct-drive, slow-



Features a marine impeller to be easy on cells.

speed stirrer (Reader Service No. 110). The Model 502D features marine impellers with variable speeds between 0 and 250 r.p.m., and a dome-based culture vessel — all designed to treat sensitive cultures gently. The 1.3 to 2.3 litre vessels are provided with a flat-bottomed jacket that allows them to stand flat on the benchtop, that can be connected to a thermocirculator for temperature control. All the normal growth parameters can be monitored with LH Fermentation's bioreactor, that can be coupled to the 500 Series modular control systems for further automation.

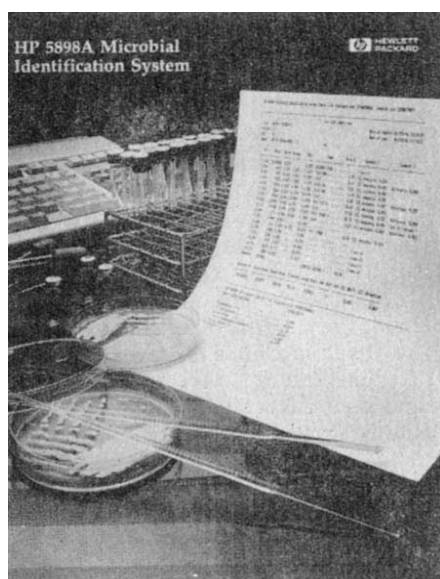
Culture conveniences

Promega announces the addition of a new six-extract package size to its Packagene **lambda DNA packaging system** (Reader Service No. 111). According to Promega, the Packagene system is a high efficiency, single-extract *in vitro* lambda packaging system that is easy to use because it eliminates unnecessary steps and packages up to 5 μg of DNA. The six-extract size \$145 (US) Packagene system also includes control lambda DNA and the LE392 plating strain.

DNASTAR Inc. has entered into a joint venture with Field Inversion Technologies Corp. to market products based upon the **field inversion gel electrophoresis** technology developed by G.F. Carle, M. Frank and M.V. Olson at Washington University (Reader Service No. 112). DNASTAR Pulse is a computer-driven controller that provides field inversion capability for up to four independent experiments simultaneously, and separates DNA molecules of up to two million base pairs. The electrophoresis programmer allows researchers to set electric fields in forward and reverse directions and to program the linear ramps optimal for the size range of given molecules. The DNASTAR Pulse system's \$1,490 (US) price includes the program, the program controller, the driver interface card, and the electrophoresis programmer with one switch module.

Throughout the year, the American Type Culture Collection, in booth number 1504, will be holding a series of workshops at its facilities in Rockville, Maryland (Reader Service No. 113). From February to June, the topics for these workshops will include freezing and freeze-drying of microorganisms, recombinant DNA, *Acanthamoeba* and *Naegleria* infections, gas chromatographic analysis of cellular fatty acids for bacterial identification, freezing and quality control of cell cultures and hybridomas, clinical anaerobic bacteriology, and biotechnology patents.

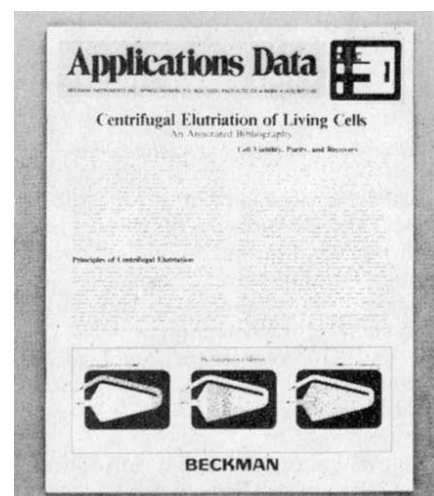
Hewlett-Packard has expanded its HP 5898A Automated Microbial Identification System to include an anaerobic database library containing more than 50 groups of anaerobic bacteria (Reader Service No. 114). This provides an automatic method for the computerized identification of microbes by means of their characteristic fatty acid profile under gas chro-



The booklet on HP's Microbial Identification System.

matography, and together with the existing database of 178 aerobic species and subspecies, makes for a complete identification system. Anaerobes included in the new database include actinomycetes, bacteroides, clostridium, fusobacterium, propionibacterium and veillonella. The library is available on floppy disk as a \$3,000 (US) option, and will be updated regularly. New libraries will be added as they become available.

To aid in the recovery and harvesting of whole living cells, Beckman has come out with a new bibliographic guide on centrifugal elutriation (Reader Service No. 115). The 24-page booklet includes more than 700 references documenting the use of Beckman elutriation systems for harvesting cells. The free booklet is divided into 27 major categories, including yeast



Tips on spinning and harvesting cells.

cells, drug suspensions and subcellular fractions. All procedures referenced in the booklet use Beckman Model JE-6B and JF-10X Elutriator Rotors. Beckman will have a full line up of its products for centrifugation in island 4E and booth numbers 422–426.

A desk-pad sized **metabolic chart** is available free of charge from Koch-Light (Reader Service No. 116). The 35 × 29 cm chart, published with the cooperation of Donald Nicholson, outlines pathways of biosynthesis and degradation of over 500 compounds. To ensure that the metabolism of carbohydrates, lipids, amino acids, and purines and pyrimidines are differentiated, and their interrelationships are made clear, the chart is printed in full colour. Koch-Light also publishes a range of charts and booklets as teaching aids, and these products are described in the current Koch-Light catalogue.

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.